PTO/SB/21 (08-03)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	10/052,664
Filing Date	January 17, 2002
First Named Inventor	Paul David Cannon
Art Unit	Nirmal Singh Basi
Examiner Name	1646
Total Number of Pages in This Submission: 55	Attorney Docket Number: ROCH-001DIV (R0058C-DIV)

ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) ____
- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☒ Appeal Communication to Group ***(Appeal Notice, Brief, Reply Brief)***
- ☐ Proprietary Information
- ☐ Status Letter
- ☒ Other Enclosure(s) *(please identify below)*:
  - Copy of Notification of Non-Compliant Appeal Brief; Certificates of Mailing; Return Postcard

Remarks: **Substitute Appeal Brief**

Applicants believe that no fees are due. However should this not be the case, the Commissioner is hereby authorized to charge any additional fees that may be required to Deposit Account No. 18-1700.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm *or* Individual name	ROCHE PALO ALTO LLC Patent Department, M/S A2-250 3431 Hillview Avenue, Palo Alto, CA 94304
Signature	David J. Chang, Reg. No. 50,374
Date	September 23, 2005

CERTIFICATE OF MAILING OR FACSIMILE

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Typed or printed name	Barbara Nolan		
Signature	Barbara Nolan	Date	September 23 , 2005

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as first class mail in an envelope addressed to:
Mail Stop: Board of Patent Appeals and Interference, US Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450, on the date indicated below.
Name of Person Mailing: Barbara Nolan

Signature: Barbara Nolan
Date: September 23, 2005

PATENT
Attorney Docket No.: ROCH-001DIV
(R0058C-DIV)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of:	Examiner: Nirmal Singh Basi
Paul David Cannon et al.	Technology Unit: 1600
Application No.: 10/052,664	Art Unit: 1646
Filed: January 17, 2002	Appeal Brief
For: HUMAN INTESTINAL NPT2B	
Assignee: Roche Palo Alto LLC	

Board of Patent Appeals and Interference
Alexandria, VA 22313-1450

APPEAL BRIEF (37 CFR §41.37)

This is an appeal from the Final Rejection in the Office Action mailed May 19, 2004, by the U.S. Patent and Trademark Office (USPTO) in the above referenced patent application. A Notice of Appeal was timely filed on July 12, 2004. This substitute Appeal Brief is filed in response to the Notification of Non-Compliant Appeal Brief sent by the USPTO on August 26, 2005 and replaces the Appeal Brief filed on October 20, 2004 which has been considered defective. Jurisdiction over this Appeal resides in the Board of Patent Appeals and Interferences (the Board) under 35 USC § 134. Applicants reserve the right to request an oral hearing.

I. Real Party in interest

Roche Palo Alto LLC, the assignee of the above referenced patent application is the real party in interest.

II. Related Appeals and Interferences

There are no related appeals or interferences.

III. Status of Claims

Claim 1 is pending in this application. Claim 1 has been rejected under 35 USC § 101 and 35 USC § 112, first paragraph. The rejection of Claim 1 is appealed.

IV. Status of Amendments

No amendments have been filed subsequent to final rejection.

V. Summary of Claimed Subject Matter

The presently-claimed invention relates to a polypeptide composition for "(a) novel human sodium phosphate co-transporter expressed in intestinal epithelial cells" (stated on page 4, lines 9-10 in Substitute Specification filed on May 10, 2002), designated as Npt2B. Using the procedures disclosed in Experimental section A (page 28, line 24 to page 29, line 21), the Npt2B polypeptide was determined to have the amino acid sequence shown in Fig. 1 and identified as SEQ ID NO:1 (stated on page 5, line 5). A description of the function of Npt2B appears on page 4, lines 19-24 in the Substitute Specification:

> "Npt2B is a type II sodium phosphate co-transporter. In its native environment, Npt2B is a co-transporter of sodium cation and phosphate anion. Npt2B is expressed, among other locations, on the surface of intestinal epithelial cells, i.e. on the apical or intestinal luminal side of the epithelial cells, and therefore

> provides for the transport of sodium and phosphate ions from the intestinal lumen into the intestinal epithelial cells."

The function of the Npt2B polypeptide as a human intestinal sodium phosphate co-transporter was derived both from the homology to published sequences for the type II intestinal transporters from *Xenopus* and mouse (page 29, lines 15-21 and references cited therein, Ishizuya-Oka et al., *Development Genetics* 20:53-66, 1997; Hilfiker et al., *Proc. Natl. Acad. Sci. U.S.A.* 95:14564-14569, 1998) **and** from the expression of the Npt2B cDNA in mammalian cells and assaying for sodium-phosphate transporter activity as described in Experimental section B (page 29, line 23 to page 30, line 19). To date, Npt2B is the sole type II human sodium phosphate co-transporter known that is found in the intestine (see Xu et al., *Biochim Biophys Acta* 1567:97-105, 2002; Werner & Kinne, *J Physiol Regul Integr Comp Physiol* 280(2):R301-312, 2001; both references cited in the Amendment and Response submitted by Applicants/Appellants on February 23, 2004).

Descriptions of the utilities of the claimed invention appear throughout the specification, and are stated, for example, on page 9, lines 28-30 as: "[t]he subject polypeptide and nucleic acid compositions find use in a variety of different applications, including research, diagnostic, and therapeutic agent screening/discovery/preparation applications, as well as in therapeutic compositions and methods employing the same." One specific utility is its use in various screening assays designed to identify therapeutic agents. As stated in the specification on page 17, lines 9-21:

> "The subject Npt2B polypeptides find use in various screening assays designed to identify therapeutic agents. Thus, one can use a cell model such as a host cell, e.g. CHO, HEK293, COS7, Xenopus Oocyte, etc., which has been transfected in a manner sufficient to express Npt2B on its surface. One can then contact the cell with a medium comprising sodium and phosphate ions, and measure the amount of phosphate anions that are internalized in the cell, where measurements are taken in both control environments and test environments, e.g. in the presence of a candidate Npt2B modulator compound, e.g. an Npt2B agonist or an Npt2B antagonist or inhibitor. To assist in detection of Pi uptake, labeled phosphorous is

> present in the medium, where any convenient label may be employed, such as an isotopic label, e.g. as present in ^{32}P or ^{33}P. Alternatively, current measurements may be taking using well known electrophysiological methods (see e.g. Electrophysiology, A practical Approach (IRL Press)(1993)), from which the uptake of Pi may be determined. Examples of assays for measuring Pi uptake are provided in: Maganin et al., *Proc. Nat'l Acad. Sci USA* (July 1993) 90: 5979-5983; and Helps et al., *Eur. J. Biochem.* (1995) 228: 927-930."

Another example of an Npt2B modulatory agent is "antibodies that at least reduce, if not inhibit the target Npt2B activity in the host. Suitable antibodies are obtained by immunizing a host animal with peptides *comprising all or a portion of the target protein, e.g. Npt2B* [emphasis added]" (Specification at page 21, lines 3-5).

The significance of identifying Npt2B modulatory agents or compounds, which either increase Npt2B activity (i.e. enhances intestinal phosphate absorption), or reduce or inhibit Npt2B activity (i.e. stops or limits intestinal phosphate absorption) appears on page 27, lines 19-29, as follows:

> "The subject methods [of modulating Npt2B activity] find use in the treatment of a variety of different disease conditions involving Npt2B activity. As such, the disease conditions treatable according to the subject methods include diseases characterized by abnormally high Pi absorption and disease conditions characterized by abnormally low Pi absorption. Disease conditions resulting from abnormally low Npt2B activity are those characterized by the presence of hypophosphatemia, and include: osteomalacia, hypocalciurea, rickets, and the like. Disease conditions resulting from abnormally high Npt2B activity are those characterized by the presence of hyperphosphatemia and include: hyperparathyroidism, hypocalcemia, vitamin D deficiency, soft tissue or metastatic calcification, and the like. Of particular interest is the use of the subject methods to treat hyperphosphatemia resulting from renal insufficiency, e.g. caused by renal disease resulting in at least impaired renal function, and the like."

Therefore, use of the Npt2B polypeptide to identify modulatory agents that increase Npt2B activity (i.e. increase intestinal phosphate absorption) would be therapeutically desirable for the treatment of diseases resulting from hypophosphatemia. Conversely,

use of the Npt2B polypeptide to identify or prepare modulatory agents that decrease Npt2B activity (i.e. decrease intestinal phosphate absorption) would be therapeutically desirable for the treatment of diseases resulting from hyperphosphatemia.

VI. Grounds of Rejection to be Reviewed on Appeal

A. Whether the invention as defined by Claim 1 is patentable under 35 USC § 101 because it has a specific and substantial utility or a well-established utility; and

B. Whether the specification enables Claim 1 under 35 USC § 112, first paragraph, since the invention is supported by a specific and substantial asserted utility or a well-established utility.

VII. Argument

A1. Utility under 35 USC § 101

In the Office Action dated November 20, 2003 ("OA 11/20/03"), the Examiner rejected claim 1 under 35 USC § 101 as allegedly being not supported by either a specific and substantial asserted utility or a well-established utility. The Examiner did not discuss the credibility of the utility asserted in the specification. The rejection of claim 1 under 35 USC § 101 was maintained in the Final Office Action dated May 19, 2004 ("OA 5/19/04").

B1. Standard of Rejection under 35 USC § 101

To properly reject a claimed invention under 35 USC § 101, the Examiner bears the burden of establishing a *prima facie* showing that the claimed invention lacks patentable utility, and needs to provide a sufficient evidentiary basis for factual assumptions relied upon in establishing the *prima facie* showing. If this initial burden is met, the burden then shifts to the applicant to provide evidence or argument to rebut the *prima facie* showing. Therefore, rejection under 35 USC § 101 for "lack of utility is a question of fact." *In re Swartz* 232 F.3d 862, 56 USPQ2d 1703 (Fed. Cir. 2000). Cases addressing the standard of rejection under 35 USC § 101 include:

In re Gaubert, 524 F.2d 1222, 1114, 187 USPQ 664, 666 (CCPA 1975) ("Accordingly, the PTO must do more than merely question operability – it must set forth factual reasons which would lead one skilled in the art to question the objective truth of the statement of operability.")

In re Oetiker, 977, F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992) ("[T]he examiner bears the initial burden, on review of the prior art or on any other ground, of presenting a *prima facie* case of unpatentability. If that burden is met, the burden of coming forward with evidence or argument shifts to the applicant ... After evidence or argument is submitted by the applicant in response, patentability is determined on the totality of the record, by a preponderance of evidence with due consideration to persuasiveness of argument ... If examination at the initial stage does not produce a *prima facie* case of unpatentability, then without more the applicant is entitled to grant of the patent.")

In re Brana, 51 F.3d 1560, 1566, 34 USPQ2d 1436, 1441 (Fed. Cir. 1995) ("Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence sufficient to convince such a person of the invention's asserted utility.")

C1. Application of Standard of Rejection under 35 USC § 101 to Claim 1

Applicants/Appellants submit that the Examiner has not met the burden of presenting a *prima facie* case that the claimed invention lacks patentable utility because he a) failed to provide any evidence or factual reasons why one skilled in the art would reasonably doubt the asserted utilities of the claimed Npt2B polypeptide and b) misinterpreted the facts in the field of the art and concerning factual statements contained in the specification.

In raising the rejection under 35 USC § 101 in OA 11/20/03 and maintaining the rejection in OA 5/19/04, the Examiner argued that "(t)he utility of claimed sodium phosphate co-transporter cannot be implicated solely from homology to known sodium phosphate co-transporter or their protein domains" (OA 11/20/03 page 6 lines 11-13, OA 5/19/04 page 4 lines 6-8). To support his position, the Examiner relied upon three generic review articles[1] on computational genomics (Bork & Koonin, *Nature Genetics* 18:313-318, 1998; Karp, *Bioinformatics* 14(9): 753-754, 1998; Bork & Eisenberg, *Current Opinion in Structural Biology* 8: 331-332, 1998) and cited excerpts from each paper in OA 11/20/03 on pages 7-8 as evidence which show that the utility of a protein cannot be implicated solely from homology and that such must be true for the Npt2B polypeptide.

The Examiner's argument is both flawed and contrary to fact for the following reasons. First, the main theme from the references cited by the Examiner is not that function **cannot be** predicted solely by homology but rather that function **may not necessarily be** predictable solely from sequence homology. I.e., in some cases, function ***is*** successfully predicted solely on the basis of homology. See, e.g., Bork & Koonin at [page #]: "Prediction of function using comparative sequence analysis is **extremely powerful** but, if not performed appropriately, **may** lead to the creation and propagation of assignment errors" [emphases added]. Second, none of the references cited by the Examiner make any mention of sodium phosphate co-transporters. The Examiner did not and could not provide an example of a case where a protein predicted by sequence homology to be a type II sodium phosphate co-transporter was, in fact, not such a protein or failed to have such activity, and Applicants/Appellants are not aware of any such case. Therefore although the Examiner has tried to use his references to argue that,

> "[t]he utility of claimed sodium phosphate co-transporter cannot be implicated solely from homology to known sodium phosphate co-transporter or their protein domains because the art does not provide teaching stating that all members of family of sodium phosphate co-transporter must have the same effects, the same

[1] These articles are "generic" in the sense that they do not discuss Npt2B in any way, nor any other sodium-phosphate co-transporter protein.

> ligands, and be involved in the same disease states, **the art discloses evidence to the contrary** [emphasis added]." (OA 11/20/03 page 6 lines 11-17, lines 18-23; OA 5/19/04 page 4 lines 6-11); and
>
> "[t]herefore, references discussed above disclose the unpredictability of assigning a function to a particular protein based on homology, **especially one that belongs to the family sodium phosphate co-transporter which have very different ligand specificity and functions** [emphasis added]." (OA 11/20/93 page 8 line 21 to page 9 line 2),

his statements have no factual support both from the references he cited and from knowledge in the field of sodium phosphate co-transporters. In contrast, there is a preponderance of factual evidence which support Applicants/Appellants' assertion that the claimed Npt2B polypeptide has utilities that are specific, substantial, and well-established, as presented later in this section C1.

As previously stated in the Summary of Claimed Subject Matter (Section V) and briefly restated here, the specification discloses a specific function for the claimed invention, which is a human type II sodium phosphate co-transporter that provides for the transport of sodium and phosphate ions from the intestinal lumen into the intestinal epithelial cells (page 4, lines 9 and 19-24). It does not require any other ligand for activity. This is a unique function for Npt2B since no other type II human ***intestinal*** sodium phosphate co-transporter have been identified (see Xu et al., *Biochim Biophys Acta* 1567:97-105, 2002; Werner & Kinne, *J Physiol Regul Integr Comp Physiol* 280(2):R301-312, 2001). Evidence confirming this function and activity of Npt2B was provided in the Declaration under 37 CFR § 1.132 by Suryananrayana Sankuratri, filed on February 23, 2004 ("Declaration"). The claimed Npt2B protein was not, at the time of filing, an "orphan" protein, i.e. a protein that has no known function, as asserted by the Examiner (see, e.g. OA 11/20/03 at page 6, line 1 and line 8; OA 5/19/04 at page 4, line 2). Therefore, the facts do not support the Examiner's statement that "[i]n light of the specification the skilled artisan can not come to any conclusions as to the function of claimed sodium phosphate co-transporter of SEQ ID NO:1" (OA 11/20/03 page 4 lines 13-15).

Based on this unique and specific function of Npt2B, the specification discloses several specific and substantial utilities for the claimed invention. One such utility for the Npt2B polypeptide is its "use in various screening assays designed to identify therapeutic agents [that modulate Npt2B activity]." (page 17, lines 9-10). This utility is "specific and substantial." It is "specific" in that using Npt2B as a screening target will identify drug candidates that specifically modulate Npt2B, rather than other proteins. This utility is "substantial" because phosphate must be transported from the intestinal lumen across the epithelium in order to be absorbed from the diet, and as Npt2B is the sole known intestinal phosphate transporter, modulation of its activity is capable of regulating phosphate absorption by the entire body. Thus, screening drug candidates for modulators of Npt2B is not merely an empty exercise, nor for academic interest alone, but one that has immediate commercial applicability.[2] Further, this utility is credible: the Examiner has presented no reason why one of ordinary skill in the art would fail to believe that Npt2B is useful as a screening target, nor are Applicants/Appellants aware of any such reason.

Applicants/Appellants' claimed protein is also independently useful as an antigen for the preparation of antibodies (see specification, page 21, lines 3-5). This utility is also specific and substantial. It is "specific" to Npt2B, because the proper generation of antibodies that specifically bind Npt2B will result in antibodies that specifically bind ***only*** to Npt2B, and not to other proteins or antigens. This utility is "substantial" because, inter alia, such antibodies can be used as inhibitors of Npt2B directly, and thus are potential drug candidates with immediate utility, as discussed above. Such antibodies are also useful for other purposes, for example detecting the expression (or absence of expression) of Npt2B, for labeling and sorting cells that express Npt2B, and the like. The Examiner

[2] Applicants/Appellants do not here assert that a particular *drug* is thus enabled: however, Applicants/Appellants point out that the right to screen a target has immediate commercial potential, e.g., for licensing to pharmaceutical

has presented no reason for doubting the utility of Npt2B used as an antigen to prepare specific antibodies, nor are Applicants/Appellants aware of any such reason: thus, this utility is also credible.

Specific diseases that are treatable by modulating Npt2B activity, either by modulatory compounds or antibodies are disclosed in the specification on page 27 lines 19-29, and were previously stated in the Summary of Claimed Subject Matter. For both utilities, the claimed Npt2B polypeptide was able to provide benefit to the public as of the filing date by identifying modulatory agents for the treatment of diseases associated with high Npt2B activity, e.g. hyperphosphatemia or with low Npt2B activity, e.g. hypophosphatemia (for list of diseases, see page 27, lines 19-29).

Statements in both OA 11/20/03 and OA 5/19/04 show that the Examiner is apparently under the misapprehension that it is necessary to correlate a disease state with a **dysfunctional** form of the claimed protein (see, e.g., OA 11/20/03 at page 5, lines 1-5; OA 5/19/04 at page 4, lines 2-4). This, however, is not necessary for the practice of Applicants/Appellants' invention. Applicants/Appellants instead find that there is utility in inhibition or stimulation of the natural, active Npt2B protein, just as many analgesics inhibit the normal function of receptors or enzymes involved in the signal chain that results in the perception of pain. Because Npt2B is uniquely situated to control absorption of phosphate from the diet, modulation of Npt2B activity is capable of affecting the amount of phosphate that is absorbed. Thus, in diseases or syndromes that are characterized by excessive (or inadequate) levels of phosphate, modulation of Npt2B activity will be therapeutic. For example, where renal failure causes hyperphosphatemia, an Npt2B inhibitor can reduce the amount of phosphate absorbed, and thus alleviate that symptom (see, e.g., Specification at page 27, lines 27-29). No dysfunction in Npt2B itself is required: one can treat conditions of too much or too little phosphate that are not

---

companies. Consider, for example, the revenues that Chiron Corp. has extracted from other companies for rights to screen various HCV proteins for inhibitors.

due to a dysfunctional Npt2B, just as one can treat pain or inflammation that are not due to a dysfunctional pain receptor.

As evidence to support the asserted function and utilities of the Npt2B protein, Applicants/Appellants submitted the Declaration of Suryanarayana Sankuratri, filed on February 23, 2004, which contained figures and data showing functional characteristics of the claimed Npt2B polypeptide **derived by following the procedure disclosed in the specification,** Experimental section B (page 29, line 23 to page 30, line 19). The Declaration presented data verifying the phosphate transporting activity claimed in the specification, which proves that the disclosure in the specification was **fully enabled** when filed. The use of the Declaration in this case is analogous to the use of the Michael Kluge declaration in *In re Brana,* 51 F.3d 1560, 1566, 34 USPQ 2d 1436, 1441 (Fed. Cir. 1995) in that the data presented was used not to **identify** or **establish** a utility but to **substantiate** utility already asserted in the specification. The relevant wording in *In re Brana* is as follows:

> "Enablement, or utility is determined as of the application filing date. *In re Glass*, 492 F.2d 1228, 1232, 181 USPQ 31, 34 (CCPA 1974). The Kluge declaration, though dated after applicants' filing date, can be used to **substantiate** any doubts as to the asserted utility since this pertains to the accuracy of a statement already in the specification. *In re Marzocchi*, 439 F.2d at 224 n.4, 169 USPQ at 370 n.4. It does not render an insufficient disclosure enabling, **but instead goes to prove that the disclosure was in fact enabling when filed** (i.e. demonstrated utility) [emphases added]." (*In re Brana*, 34 USPQ2d at 1441)

In OA 5/19/04, the Examiner sought to disparage the evidence presented in the Declaration by stating "[i]n instant case post filing art cannot be used to **establish** utility because the results of said art were not known at the time of filing of instant application, and the information obtained was due to further experimentation [emphasis added]." (page 4 lines 18-21) The Examiner's position is contrary both to the facts of the present case and to case law.

Further evidence on the asserted utilities of human Npt2B in the specification is ascertained from the general knowledge in the area of Type II sodium phosphate cotransporters, especially the Type IIb intestinal transporter at around the time of filing of the priority application (February 9, 1999). Hilfiker et al. (*Proc. Natl. Acad. Sci. U.S.A.* 95:14564-14569, 1998, cited in the specification on page 2, lines 17-18 and page 29 lines 20-21), published in November, 1998, described the cloning and characterization of mouse Npt2B and was the first paper to classify the Type II sodium phosphate co-transporters into the "Type IIa" family, represented by the renal isoforms and the "Type IIb" family, represented by the intestinal isoforms, which includes Npt2B. Xu et al. (*Genomics* 62:281-284, 1999, cited by the previous Examiner in this case in the Notice of Allowance sent on September 11, 2003) describe the cloning and characterization of human Npt2B (referred in the paper as Na/Pi-IIb), and was published ten months after Applicants/Appellants' priority date. However, case law has stated that the "court has approved use of later publications as evidence of the state of art *existing on the filing date* of an application." *In re Hogan and Banks* 559 F.2d 595, 605, 194 USPQ 527, 537 (CCPA 1977). Both papers describe their proteins (which have 78.8% and 99.9% sequence identity to the Npt2B polypeptide of Claim 1) as the intestinal or type IIb sodium phosphate co-transporter. The significance attributed by the authors to their findings can be represented by the following statement in the Abstract from Xu et al.:

> "Phosphate plays a crucial role in cellular metabolism and its homeostatic regulation in intestinal and renal epithelial is critical. Apically expressed sodium-phosphate (Na^+-Pi) transporters play a ***critical*** role in this regulation. We have isolated a cDNA (HGMW-approved symbol SLC34A2) encoding a novel human small intestinal Na^+-Pi transporter [emphasis added]."

Therefore, a person of ordinary skill in the art would, at the time of filing, immediately appreciate why the claimed Npt2B polypeptide is useful, based on the characteristics of the invention. Applicants/Appellants also submit that based on the evidence presented concerning the knowledge possessed by one skilled in the art at the time of filing, the

asserted specific and substantial utility for the claimed invention also qualifies as a well-established utility.

Current knowledge in the area of Type IIb intestinal sodium phosphate transporters also fully contradict the Examiner's assertions that the claimed Npt2B polypeptide has no specific or substantial utility. As set forth in the specification at page 29 lines 14-15, Applicants/Appellants confirmed by RT-PCR that Npt2B is expressed in the small intestine. To date, Npt2B is the sole Type II human sodium phosphate co-transporter known that is found in the small intestine (see Xu et al., *Biochim Biophys Acta* 1567:97-105, 2002; Werner & Kinne, *J Physiol Regul Integr Comp Physiol* 280(2):R301-312, 2001; both references cited in the Amendment and Response submitted by Applicants/Appellants on February 23, 2004). This **fact** demonstrates the unique function of Npt2B and that its utility does not rest solely on homology arguments, as asserted by the Examiner throughout both OA 11/20/03 and OA 5/19/04.

Furthermore, a recent article by Peerce et al. (*Biochem. Biophys. Res. Commun.* 301:8-12, 2003), cited in the Declaration of Suryanarayna Sankuratri, states that "[a] pharmacological method of reducing intestinal phosphate absorption may provide a more palatable approach to reducing serum phosphate and may slow the progression of moderate chronic renal failure to end-stage renal failure. In the proximal small intestine phosphate absorption occurs by a Na^+-dependent mechanism ... [which] occurs through the Na^+/phosphate cotransporter. The Na^+/phosphate cotransporter has been identified as a 110-120kDa polypeptide (references including Hilfiker et al. and Xu et al.)" This statement in Peerce et al. supports and confirms the disclosure in the specification on page 27, lines 27-29 which reads, "[o]f particular interest is the use of the subject methods [of modulating Npt2B activity] to treat hyperphosphatemia resulting from renal insufficiency, e.g. caused by renal disease resulting in at least impaired renal function, and the like."

In OA 11/20/03, the Examiner sought to support the rejection under 35 USC § 101 by implying that the disclosure in the present specification was analogous to the situations decided by the Courts in *Brenner v Manson*, 383 U.S. 519, 148 USPQ 689 (1966) (page 11 lines 12-14; page 14 lines 7-12) and in *In re Kirk*, 376 F.2d 936, 153 USPQ 48 (CCPA 1967) (page 14 line 21 to page 15 line 7). These analogies are fallacious for the following reasons. In *Brenner*, the applicant failed to disclose **any** utility for a process to synthesize a steroid compound with no known utility, other than as "a possible object of scientific inquiry", and offered as evidence only a third party article showing the utility of an homologue of the subject steroid compound. In *Kirk*, the applicants claimed steroid compounds said to have valuable "biological properties" and to be of value to the furtherance of steroidal research. In contrast, the present specification discloses a specific function of the Npt2B polypeptide and its use to identify agents to treat specific diseases. Furthermore, the specification asserts both specific and substantial utilities for the Npt2B polypeptide that are supported by both the prior art and present knowledge in the field.

In summary, Applicants identified and sequenced Npt2B, determined its biological function, and set forth several utilities in the specification, including the use in screening assays and the use to generate antibodies. These utilities are unique to Npt2B, as it is the only sodium-phosphate co-transporter found in the human intestine, and therefore mediates all absorption of phosphate from the diet. Even the Examiner found that the sodium-phosphate transporter family was diverse, (see, e.g. OA 11/20/03 at page 4, line 1, and page 11, lines 17-19) and that one would not expect compounds that affect one transporter to modulate another. The Declaration of Suryanarayna Sankuratri (Appendix, __) followed the procedures set forth in the specification and confirmed the screening utility set forth therein. The Examiner, in contrast, has provided no reason for doubting the asserted utility, and has provided no factual basis for believing that any of Applicants/Appellants' utilities would not be substantial and specific to the claimed invention.

Based on the arguments set forth, Applicants/Appellants submit that under the Standard of Rejection under 35 USC § 101, the Examiner has not met the burden of presenting a *prima facie* case that the claimed invention lacks patentable utility by providing evidence showing that one of ordinary skill in the art would reasonably doubt the utility asserted in the specification. Even if the Board considers that the Examiner has met this initial burden, Applicants/Appellants submit that sufficient rebuttal evidence has been provided to convince one of skill in the art of that the asserted utility was specific, substantial, and well-established such that the utility requirement under 35 USC § 101 has been satisfied.

A2. Enablement under 35 USC § 112, first paragraph

In OA 11/20/03, the Examiner rejected claim 1 under 35 USC § 112, first paragraph as allegedly not enabling one skilled in the art how to use the claimed invention "since the invention is not supported by either a specific and substantial asserted utility or a well-established utility." The rejection of Claim 1 under 35 USC § 112, first paragraph (enablement) was maintained in OA 5/19/04.

B2. Standard of Rejection under 35 USC § 112, first paragraph (enablement)

Section 112, first paragraph, requires that the specification teach one of ordinary skill in the art how to make and use the invention. As stated by the court in *In re Marzocchi* (CCPA 1971) 439 F.2d 220, 169 USPQ 367:

> "[A] specification disclosure which contains a teaching of the manner and process of making and using the invention in terms which correspond in scope to those used in describing and defining the subject matter sought to be patented must be taken as in compliance with the enablement requirement of the first paragraph of § 112 unless there is reason to doubt the objective truth of the statements contained therein which must be relied on for enabling support." (439 F.2d at 223, 169 USPQ at 369.)

C2. Application of Standard of Rejection under 35 USC § 112, first paragraph (enablement) to Claim 1

The Examiner's rejection under 35 USC § 112, first paragraph was based solely on the utility rejection under 35 USC § 101. Reiterating the arguments set forth in Section VII C1 above, Applicants/Appellants submit that the Examiner did not meet the burden of presenting a *prima facie* case that the claimed invention lacks patentable utility by providing evidence showing that one of ordinary skill in the art would reasonably doubt the utility asserted in the specification. Therefore Claim 1 satisfies the utility requirement of 35 USC § 101.

Applicants/Appellants also assert that Claim 1 satisfies the enablement requirement of 35 USC § 112, first paragraph since the specification would have taught one of ordinary skill in the art how to make and use the invention at the time of filing. The amino acid sequence of the Npt2B polypeptide, as disclosed in Figure 1 and in SEQ ID NO:1, enabled the artisan to generate antibodies which can modulate the activity of Npt2B to treat specific diseases of phosphate metabolism. Methods of preparing such antibodies are described in the specification starting from page 21, line 3 to page 23, line 12. Also, following the procedures described in the specification in Experimental Section B (page 29, line 23 to page 30, line 19), would enable the artisan to express the Npt2B polypeptide and perform screening assays to identify Npt2B modulatory agents useful for the treatment of diseases of phosphate metabolism. Furthermore, the figures and data contained in the Declaration of Suryanarayana Sankuratri, which were obtained by following the procedures in Experimental Section B and which confirmed both the asserted activity and utility of the claimed invention, both demonstrate and prove that the specification was enabling at the time of filing.

D. Conclusion

The Examiner has failed to establish a *prima facie* showing that the asserted utility for Claim 1 is not specific or substantial. Applicants/Appellants have provided evidence

showing that the asserted utility for the present invention was specific, substantial and well-established at the time of filing. Applicants/Appellants have also shown that the specification was fully enabling at the time of filing. Applicants/Appellants have therefore demonstrated that Claim 1 satisfies the utility requirement of 35 USC § 101 and the enablement requirement of 35 USC § 112, first paragraph. Accordingly, Applicants/Appellants request that the Board of Patent Appeals and Interferences reverse the rejection of Claim 1 on the grounds set forth herein.

Respectfully submitted,

David J. Chang, Ph.D.
Reg. No. 50,374

Roche Palo Alto LLC
Patent Law Dept. M/S A2-250
3431 Hillview Avenue
Palo Alto, CA 94304

Direct Phone: (650) 855-5316
Facsimile: (650) 855-5322
Date: September 23, 2005

VIII. Claims Appendix

This Appendix contains the Claim involved in the Appeal:

1. An isolated Npt2B polypeptide comprising the amino acid set forth in SEQ ID NO:1.

IX. Evidence Appendix

This Appendix contains copies of:

A. Declaration of Suryanarayana Sankuratri under 37 C.F.R. § 1.132 filed on February 23, 2004.

B. Hilfiker et al., "Characterization of a murine type II sodium-phosphate cotransporter expressed in mammalian small intestine", *Proc. Natl. Acad. Sci. U.S.A.* 95: 14564-14569, 1998.
This reference was entered by Examiner Peter Paras in the "Information Disclosure Citation" in the Notice of Allowance and Fee Due sent on September 11, 2003.

C. Xu et al., "Molecular cloning, functional characterization, tissue distribution, and chromosomal localization of a human, small intestinal sodium-phosphate (Na+-Pi) transporter (SLC34A2)" *Genomics* 62: 281-284, 1999.
This reference was cited by Examiner Peter Paras in the "Notice of References Cited" in the Notice of Allowance and Fee Due sent on September 11, 2003.

D. Bork & Koonin, "Predicting functions from protein sequences – where are the bottlenecks?" *Nature Genetics* 18: 313-318, 1998.
This reference was relied upon by the examiner as to grounds of rejection in the Office Action of November 20, 2003.

E. Karp, "What we do not know about sequence analysis and sequence database" *Bioinformatics* 14(9): 753-754, 1998.
This reference was relied upon by the examiner as to grounds of rejection in the Office Action of November 20, 2003.

F. Bork & Eisenberg, "Sequences and topology- deriving biological knowledge from genomic sequences" *Current Opinion in Structural Biology* 8:331-332, 1998.

This reference was relied upon by the examiner as to grounds of rejection in the Office Action of November 20, 2003.

IX A.

Atty Dkt. No.: R0058C-DIV
USSN: 10/052,664
Exhibit 1

CERTIFICATE OF MAILING		
I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.		
Typed or Printed Name		
Signature Dais Weise	**Date** 2-18-04	
DECLARATION OF SURYANARAYANA SANKURATRI UNDER 37 C.F.R. § 1.132 Address to: Assistant Commissioner for Patents Washington, D.C. 20231	Attorney Docket Confirmation No.	ROCH-001DIV 4008
	First Named Inventor	Paul David Cannon
	Application Number	10/052,664
	Filing Date	January 17, 2002
	Group Art Unit	1646
	Examiner Name	Nirmal Singh Basi
	Title	*Human Intestinal Npt2B*

I, Suryanarayana Sankuratri, do hereby declare and state:

1. I am a biochemist currently employed as Principal Research Scientist at Roche Palo Alto, LLC, Palo Alto, California, and am a co-inventor of the claims of the above-identified patent application. I directed others and personally performed the research leading to the invention disclosed and claimed therein. My professional experience, educational background, professional activities, and publications are detailed in the curriculum vitae attached hereto.

2. I have read the Office Action dated November 20, 2003, in this application and understand that the Examiner has rejected pending Claim 1 on the assertion that the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility. This assertion is incorrect for the reasons set forth below.

3. Using the procedure that was disclosed in the application (page 29 line 25 to page 30 line 19), we were able to show that CHO cells which express the human Npt2B protein of this invention were able to transport phosphate ions as measured by the amount of radioactive phosphate taken up by the cells, whereas CHO cells not expressing Npt2B did not transport phosphate. This result, graphically represented in Fig. 1, clearly demonstrated that Npt2B is a phosphate transporter.

4. We then compared the biochemical characteristics of Npt2B with those of the renal phosphate transporter, Npt2A. As seen in Fig. 2, both transporters required the presence of

sodium in order to transport phosphate. Fig. 3 compares the phosphate uptake kinetics of the two transporters, showing that Npt2B demonstrated higher affinities for both sodium and phosphate ions than Npt2A. In fact, Km measurements for sodium and phosphate uptake for Npt2B were remarkably similar to those obtained from intact intestinal membrane vesicles (Peerce BE, *Biochim Biophys Acta.* 1239: 1-10, 1995).

5. We also compared the pH dependence of phosphate transport between Npt2A and Npt2B. As seen in Figure 4, the two transporters had opposite responses to pH changes, with Npt2B showing decreased phosphate uptake as the assay conditions shifted from acidic to alkaline whereas Npt2A showed increased phosphate uptake as pH increased. This pH dependence is one of the characteristic features of intestinal phosphate transporter characterized by many laboratories using intestinal membrane preparations.

6. We conducted a Northern blot analysis (Fig. 5) which clearly showed that Npt2B was expressed in the ileum, jejunum and duodenum but Npt2A could not be detected in these areas of the intestine. It is now well-established in the scientific community that Npt2B is the protein involved in intestinal sodium-dependent phosphate absorption. It is also well-accepted by researchers in the field that complications such as hyperphosphatemia which could both cause and be caused by renal disease can be treated by reducing the amount of phosphate absorption from the intestine. A recent article by Peerce et al. (*Biochem Biophys Res Commun.* 301: 8-12, 2003, attached herein) states, "A pharmacological method of reducing intestinal phosphate absorption may provide a more palatable approach to reducing serum phosphate and may slow the progression of moderate chronic renal failure to end-stage renal failure." As Npt2B is responsible for most of the phosphate absorbed from the diet, it is uniquely situated to control the amount of phosphate that enters the system. Therefore, the use of Npt2B in a screening assay to identify inhibitors of the transporter would be of significant importance.

7. Using CHO cells expressing Npt2B, we were successful in identifying a number of compounds which were able to inhibit Npt2B activity at low micromolar concentrations (Fig. 6). These and other newly identified inhibitors of Npt2B could play a significant role in the treatment of diseases characterized by hyperphosphatemia.

8. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title XVIII of the United States Code, and that such will false statements may jeopardize the validity of the application or any patent issuing thereon.

2/13/04
Date

Suryanarayana Sankuratri

Fig. 1 Heterologous expression of the novel phosphate transporter in mammalian cells

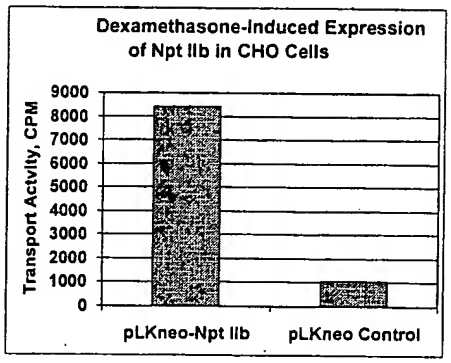


Fig. 2 Na/Pi Cotransport Activity of Npt IIa & Novel Transporter Npt IIb In CHO Cells

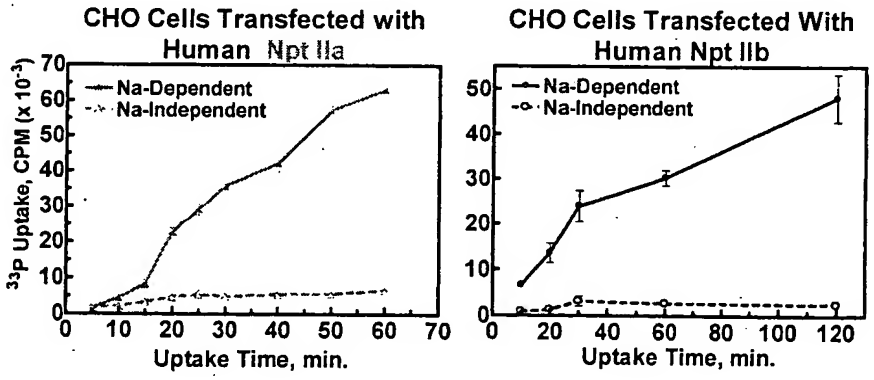
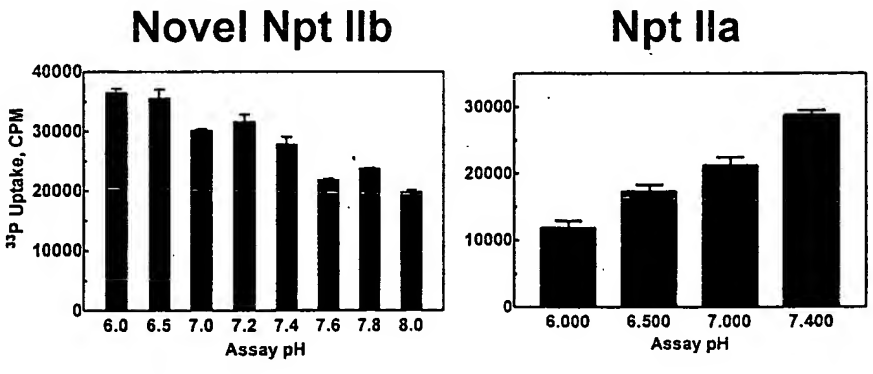


Fig. 3 Equilibrium constant (Km) for the novel phosphate transporter human Npt IIb and human renal Npt IIa

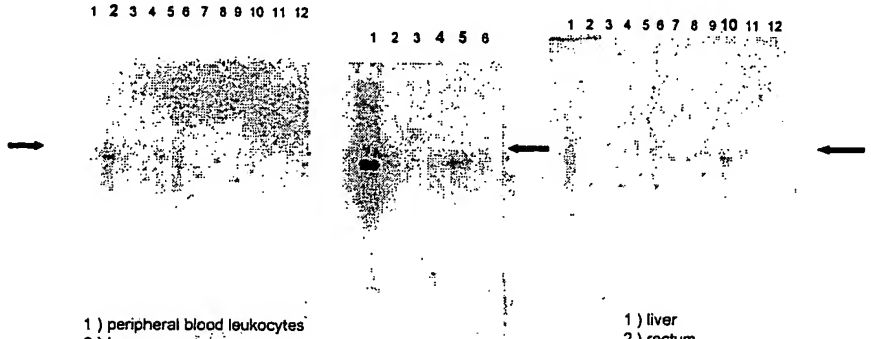
	Novel Npt IIb	Npt IIa
Phosphate (μM)	50	80
Sodium (mM)	30	128

Fig. 4 pH-dependent activity of novel transporter



Expressed in CHO cells

Fig. 5 Northern Analysis for new transporter expression



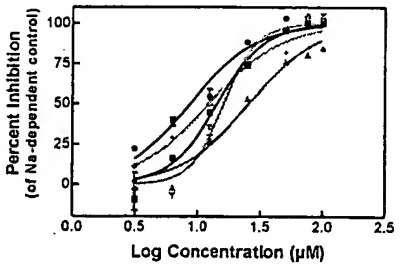
1) peripheral blood leukocytes
2) lung
3) placenta
4) small intestine
5) liver
6) kidney
7) spleen
8) thymus
9) colon
10) skeletal muscle
11) heart
12) brain

1) lung
2) rectum
3) colon
4) ileum
5) jejunum
6) stomach

1) liver
2) rectum
3) cecum
4) transverse colon
5) descending colon
6) ascending colon
7) jejunum
8) ileum
9) ileocecum
10) duodenum
11) stomach
12) esophagus

Fig. 6 Npt IIb selective inhibitors

Data Table-6



- RO-47-3128/000 #142
- RO-48-0667/000
- RO116-9033/000
- RO116-9580/003 #266
- RO116-4333/003

	RO-47-3128/000	RO-48-0667/000	RO116-9033/000	RO116-9580/003	RO116-4333/003
EC50	14.36	25.60	16.23	12.33	8.643

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 301 (2003) 8–12
BBRC
www.elsevier.com/locate/ybbrc

Inhibition of human intestinal brush border membrane vesicle Na^+-dependent phosphate uptake by phosphophloretin derivatives

Brian E. Peerce,[a,*] R.Y. Declan Fleming,[b] and Rebecca D. Clarke[a]

[a] *Department of Physiology and Biophysics, University of Texas Medical Branch, 12th and Mechanic, 2.200 Basic Science Bldg., Galveston, TX 77555-0641, USA*
[b] *Department of Surgery, University of Texas Medical Branch, Galveston, TX 77555-0641, USA*

Received 9 December 2002

Abstract

Hyperphosphatemia and II° hyperparathyroidism are common and severe complications of chronic renal failure. Reduced dietary phosphorus has been shown to be an effective treatment in reducing serum phosphate and serum PTH. 2′-Phosphophloretin inhibited small intestine apical membrane Na^+/phosphate cotransport and reduced serum phosphate in adult rats. 2′-PP and phosphoesters of phloretin were tested for inhibition of human small intestine brush border membrane alkaline phosphatase activity and for inhibition of Na^+-dependent phosphate uptake. The IC_{50}'s for inhibition of alkaline phosphatase suggested an order of inhibitory potency of 4-PP > phloretin > 4′-PP > 2′-PP. Inhibition of Na^+-dependent phosphate uptake followed the sequence 2′-PP ≫ 4′-PP > 4-PP > phloretin. These results are consistent with 2′-PP being a specific inhibitor of human intestinal brush border membrane Na^+/phosphate cotransport.

Keywords: Human intestinal brush border membrane; Intestinal phosphate absorption; Na^+-dependent phosphate uptake; Na^+/phosphate cotransporter; 2′-Phosphophloretin

In chronic renal failure phosphate retention and deposition as calcium phosphate precipitates contribute to interstitial injury, renal tubule injury, and cardiac disease [1–4]. Very low phosphorus diets in combination with phosphate binding compounds have been shown to slow the progression of renal failure. A pharmacological method of reducing intestinal phosphate absorption may provide a more palatable approach to reducing serum phosphate and may slow the progression of moderate chronic renal failure to end-stage renal failure.

In the proximal small intestine phosphate absorption occurs by a Na^+-dependent mechanism and a Na^+-independent process. Na^+-dependent phosphate uptake occurs through the Na^+/phosphate cotransporter. The Na^+/phosphate cotransporter has been identified as a 110–120 kDa polypeptide [5–8]. The mechanism of Na^+-independent uptake is unknown.

A phosphate ester of phloretin has been shown to inhibit rat and rabbit intestinal brush border membrane vesicle Na^+-dependent phosphate uptake [9]. 2′-PP inhibition of brush border membrane (BBM) Na^+-dependent phosphate uptake required Na^+ and was sensitive to external phosphate. In vivo 2′-PP reduced plasma phosphate in rats in a concentration-dependent manner. We have extended our studies of the effect of phosphophloretins to human BBM alkaline phosphatase activity and phosphate uptake into human BBM vesicles.

Materials and methods

Materials. Chemicals used in the synthesis of 2′-PP, 4′-PP, and 4-PP were purchased from Aldrich Chemical, Milwaukee, WI. 3-(4-hydroxyphenyl)-propionitrile was purchased from Lancaster Chemical, Lancaster, PA. All organic solvents were purchased from Aldrich Chemical, Milwaukee, WI and were of reagent grade or better. Membrane filters were purchased from Millipore, Boston, MA. [^{32}P]Phosphate was purchased from DuPont/NEN, Wilmington, DE. Salts and

* Corresponding author. Fax: 1-409-772-3381.
E-mail address: bpeerce@UTMB.edu (B.E. Peerce).

0006-291X/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0006-291X(02)02974-1

reagents used in the preparation and assay of brush border membrane vesicles were purchased from Fisher Chemical, Houston, TX.

Methods

Preparation of brush border membrane vesicles. Human intestine removed during surgical procedures was scraped and the mucosa was stored in 300 mM mannitol and 10 mM Hepes/Tris, pH 7.5, at liquid N_2 temperatures until needed. Brush border membrane vesicles were prepared by Ca^{2+} precipitation and differential centrifugation as previously described [5,6,10–15]. Purification of brush border membranes was assayed using the brush border membrane enzyme markers sucrase [16] and alkaline phosphatase [17]. During the course of these studies, enrichment in brush border membrane enzymes varied between 20- and 28-fold.

Synthesis of phosphophloretin derivatives. 2′-Phosphophloretin (2′-PP) was synthesized from phloridzin [9]. 2′-PP was analyzed by Mass Spectrometry, ^{31}P NMR ^{13}CNMR, and ^{1}HNMR [9]. ^{1}H NMR (400 Hz, d_6-DMSO) δ 13.0 (s, 1H), 10.7 (br. s, 1H), 9.2 (br. s., 1H), 7.03 (d, $J = 8.6$ Hz, 2H), 6.64 (d, $J = 8.4$ Hz, 2H), 6.63 (dd, $J = 1.2$, 2.1, 1H), 6.04 (d, $J = 2.4$ Hz, 1H), 3.27 (t, $J = 7.2$ Hz, 2H), 2.77 (t, $J = 7.6$ Hz, 2H); ^{31}P NMR δ-4; ESMS *m/z* 355 (M + H); and melting point = 170–171 °C.

4′-Phosphophloretin (4′-PP) was synthesized from 2,6-dihydroxy-4-phospho benzene and 4-hydroxy phenyl propionyl nitrile [18]. The 4′-phosphoester was resolved from the 2′-phosphoester by chromatography on silica gel using hexanes: dichloromethane: ethyl acetate (50:25:25). 2,6-Dihydro-4-phospho benzene was synthesized from phloroglucinol and dibenzyl phosphite in acetonitrile and triethylamine [19]. Prior to reaction with dibenzyl phosphite, phloroglucinol was dried at 105 °C under vacuum for 7 days. 2,6-Dihydro-4-phospho benzene was isolated by column chromatography on Dowex 1 using 25% methanol to elute the column. 4′-Phosphophloretin was purified by silica gel column chromatography developed with hexanes: dichloromethane: ethyl acetate (60:25:15). 4′-Phosphophloretin was analyzed by NMR and mass spectrometry. ^{1}H NMR (750 Hz, d_6 DMSO) δ 13.5 (s, 1H), 9 (br. s, 1H), 7.08 (d, $J = 8.2$ Hz, 2H), 7.06 (d, $J = 8.2$ Hz, 2H), 6.74 (s, 2H), 6.65 (d, $J = 8.2$ Hz, 2H), 6.62 (d, $J = 8.2$ Hz, 2H), 2.7 (t, $J = 7.5$ Hz, 5.1 Hz, 2H), 1.22 (s, 2H); ^{31}P NMR δ-4.8; ESMS *m/z* 355 (M + H); and melting point 178–179 °C.

4-Phosphophloretin (4-PP) was synthesized from 3-(4-dibenzyl phosphophenyl) propionyl chloride and phloroglucinol by Friedel-Crafts acylation in DMSO with anhydrous $AlCl_3$ [9,18]. The carboxylic acid of 3-(4-hydroxy)-cinnamic acid (5 g) was reacted with benzyl bromide in HMPT (hexamethylphosphoric triamide) for 1 h at 23 °C. The benzoate was collected and recrystallized from ethanol. The benzoate (5.04 g, 20 mmol) was added to 50 ml *n,n*-dimethylacetamide and cooled to 4 °C with stirring, and NaH was added (0.64 g, 25 mmol). The mixture was brought to 23 °C and 10 ml CCl_4 was added. Dibenzyl phosphite (5.6 g, 25.8 mmoles) in 25 ml *n,n*-dimethylacetamide was added and stirring was continued for 1 h at 23 °C. The reactants were diluted with 0.2 M acetate buffer, pH 4 (200 ml) and the di-benzyl phosphate ester was partitioned between water:hexane:ethyl acetate (50:25:25). The di-benzyl phosphate ester was reduced in volume, purified by chromatography on a silica gel column eluted with a 25–50% ethyl acetate gradient in hexanes, and dried at 75 °C under vacuum. The benzyl protecting groups were cleaved by catalytic hydrogenation with H_2 gas in ethyl acetate (100 ml) and 200 mg Pd/C for 24 h. 4-PP was purified as previously described [9]. 3-(4-phosphophenyl) propionyl chloride was synthesized from 3-(4-hydroxy) cinnamic acid and dibenzyl phosphite [19]. ^{1}HNMR (400 Hz, d_6 DMSO) δ 10.5 (br. s, 1H), 9.2 (br. s, 2H), 7.02 (d, 2H, $J = 8.2$ Hz), 6.8 (d, 2H, $J = 8.2$ Hz), 6.64 (d, 2H, $J = 8.4$ Hz), 6.6 (dd, $J = 2.5$, 1.5 Hz, 1H), 6.04 (d, $J = 2.5$ Hz, 1H), 3.3 (t, $J = 7.2$ Hz, 2H), 2.7 (t, $J = 7.5$ Hz); ^{31}P NMR δ-4.8: ESMS *m/z* 355 (M + H); and melting point 182 °C.

Phosphorylated phloretin derivatives were analyzed by thin layer chromatography using silica gel and methanol:H_2O (1:3) as the developing solvent. Spots were identified by UV absorption, I_2 [20] and visualized for phosphate esters using Hanes reagent [21]. Phosphophloretin derivatives were single spots and judged to be 90–94% of the UV absorbing material.

Na^+-dependent brush border membrane vesicle uptakes. Na^+-gradient driven uptakes of phosphate, alanine, or glucose into intestinal brush border membrane vesicles were performed using a rapid mixing rapid filtering device as previously described [5,6,9–15]. Na^+-dependent phosphate uptake into brush border membrane vesicles was performed using 100 μM [^{32}P] phosphate, 100 mM mannitol, 10 mM Hepes/Tris, pH 7.5, 100 mM NaCl or 100 mM KCl (uptake buffers). Na^+-dependent glucose uptake was determined using 100 μM [^{3}H] glucose, 10 mM Hepes/Tris, pH 7.5, 100 mM mannitol, and 100 mM NaCl or 100 mM KCl. Na^+-dependent alanine uptake was determined using 100 μM [^{3}H] alanine, 100 mM mannitol, 10 mM Hepes/Tris, pH 7.5, and 100 mM NaCl or 100 mM KCl. Uptakes were performed at 23 °C using 100 μg of brush border membrane protein.

Experiments examining the effect of phosphophloretin derivatives on Na^+-dependent uptakes were performed as described above using 10 nM to 10 μM phosphophloretin dissolved in 10 mM KOH:borate, pH 6.5. Phosphophloretin was added to the uptake solution immediately prior to addition of protein. In some experiments the effect of external phosphate on phosphophloretin inhibition of Na^+-dependent phosphate uptake was examined. In these experiments, phosphate concentration was varied between 25 and 500 μM. The effect of phosphate concentration on phosphophloretin inhibition of Na^+-dependent [^{32}P]phosphate uptake into intestinal brush border membrane vesicles was analyzed using the non-linear regression program, Enzfitter, Elsevier, Biosoft, Cambridge, UK.

In some experiments the time course of phosphate uptake into human intestinal BBMV was examined. Uptake of phosphate into BBMV was determined between 3 s and 30 min at 23 °C. Na^+-dependent uptakes were defined as uptake in the presence of NaCl minus uptake in the presence of KCl. All uptakes were performed in triplicate and the results are expressed as means ± SE.

Measurement of BBM alkaline phosphatase activity. Intestinal BBM alkaline phosphatase activity was measured using 1 mM *p*-nitrophenylphosphate and 100 μg BBM protein as previously described [17]. In experiments examining the effect of phosphophloretin derivatives on alkaline phosphatase activity the indicated phosphophloretin derivative was varied between 100 nM and 100 μM.

Results

Effect of phosphophloretins on Na^+-dependent phosphate uptake

The time course of phosphate uptake into human intestinal BBMV is shown in Fig. 1. Phosphate uptakes into BBMV in the presence of NaCl (closed circles, solid line), in the presence of KCl (open squares, dashed line), and in the presence of NaCl and 100 nM 2′-PP (open circles, solid line) are shown. Fig. 1 shows a 7-fold overshoot for phosphate uptake over equilibrium phosphate uptake in the presence of NaCl. Addition of 100 nM 2′-PP resulted in a 75–80% decrease in phosphate uptake without affecting phosphate uptake at equilibrium. During the course of these studies, the phosphate overshoot of equilibrium phosphate accumulation varied between 5- and 12-fold (mean = 7.8-fold, $n = 5$).

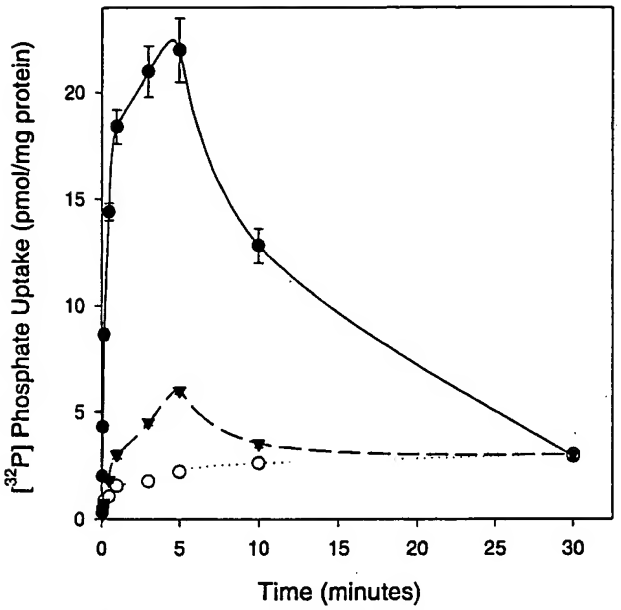


Fig. 1. Time course of phosphate uptake into human small intestinal BBMV. [^{32}P]phosphate uptake into human small intestine BBMV was determined as described in Materials and methods. Phosphate uptakes in the presence of NaCl (closed circles, solid line), in the presence of KCl (open circles, dotted line), and in the presence of NaCl + 100 nM of 2′-PP (inverse triangles, dashed line) were determined following 3 s to 30 min incubations at 23 °C. Results are means ± SE of triplicate determinations and representative of five experiments.

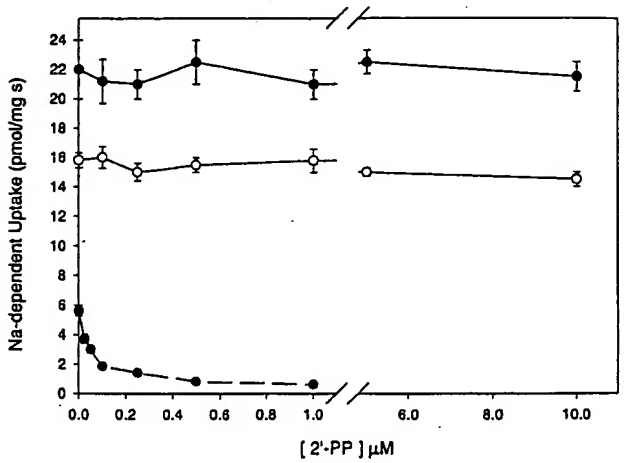
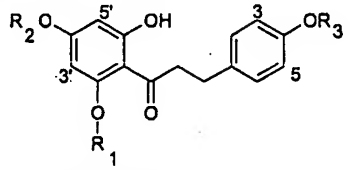


Fig. 2. Effect of 2′-PP on Na^+-dependent Cotransport Na^+ dependent [^{32}P]phosphate (solid circles, dashed line), Na^+-dependent [^{3}H] glucose (solid circles, solid line), or Na^+-dependent [^{3}H] alanine (open circles, solid line) uptakes were determined as described in Materials and methods. Results are means ± SE of triplicate determinations and representative of three separate experiments.

Fig. 2 shows the effect of 2′-phosphophloretin concentration on Na^+-dependent transport into BBM vesicles. 2′-PP inhibited Na^+-dependent phosphate uptake (solid circles, broken line) in a concentration-dependent manner with an apparent IC_{50} of 38 ± 6 nM ($n = 4$). Na^+-dependent glucose uptake (solid circles, solid line) and Na^+-dependent alanine uptake (open circles, solid line) were not affected by 2′-PP at concentrations 10 times that required for greater than 90% inhibition of Na^+-dependent phosphate uptake.

Studies examining the effect of phosphophloretins on Na^+-dependent phosphate uptake and alkaline phosphatase activity are summarized in Table 1. Na^+-dependent phosphate uptake was insensitive 4-PP and

Table 1
Effect of phosphorylated aromatics on Na-dependent phosphate uptake



Compound	Na^+-dependent phosphate uptake		Alkaline phosphatase activity
	IC_{50} (μM)	% Change	IC_{50} (mM)
2′-PP		Inhibition	
$R_2 = R_3 = H$	0.038 ± 0.006	92 ± 4	1.25 ± 0.25
$R_1 = HPO_4$			
4′-PP		Inhibition	
$R_1 = R_3 = H$	NM	15 + 4	0.96 ± 0.08
$R_2 = HPO_4$			
4-PP		Stimulation	
$R_1 = R_2 = H$	0.185 + 0.02	38 ± 12	0.350 ± 0.08
$R_3 = HPO_4$			
Phloretin	NM		0.692 ± 0.058

NM, not measurable. Results are means ± SE of triplicate determinations and three separate experiments. *p*-nitrophenyl phosphate concentration was 1 mM.

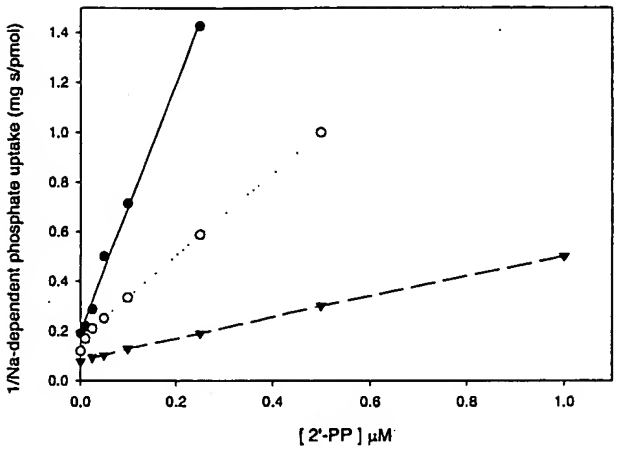


Fig. 3. Effect of [Phosphate] on 2′-PP Inhibition of Na^+-dependent phosphate uptake Na^+-dependent [^{32}P]phosphate uptake into intestinal brush border membrane vesicles was determined as described in Materials and methods. External phosphate concentration was 50 μM (open circles), 100 μM (closed circles), or 250 μM (solid triangles). 2′-PP concentration was varied between 10 nM and 1 μM. Results are means ± SE of triplicate determinations and representative of three experiments.

phloretin at concentrations below 100 μM. Addition of 4′-PP resulted in a 15% ± 4% ($n = 3$) inhibition of Na^+-dependent phosphate uptake at 500 nM 4′-PP. In contrast, Na^+-dependent phosphate uptake was inhibited more than 90% at 2′-PP concentrations above 100 nM. All of the phloretin derivatives examined were weak inhibitors of intestinal BBM alkaline phosphatase activity.

Effect of external phosphate concentration on 2′-PP inhibition of Na^+-dependent phosphate uptake

The effect of external phosphate concentration on 2′-PP inhibition of Na^+-dependent phosphate uptake is shown in Fig. 3. Fig. 3 is a Dixon plot of the effect of 50 μM phosphate (solid circles), 100 μM phosphate (open circles), and 250 μM phosphate (solid triangles) on 2′-PP inhibition of Na^+-dependent phosphate uptake. Increasing the external phosphate concentration decreased 2′-PP inhibition of Na^+-dependent phosphate uptake. The effect of phosphate concentration on 2′-PP inhibition of brush border Na^+-dependent phosphate uptake was analyzed by the method of Cornish-Bowden [22] at 50, 100, and 250 μM of 2′-PP. The intercept of the three straight lines was above the *X*-axis and to the right of the *Y*-axis, which is consistent with mixed inhibition by 2′-PP [23].

Discussion

2′-PP inhibition of Na^+-dependent phosphate uptake was measured in proximal small intestine brush border membrane vesicles isolated from human small intestine. The apparent IC_{50} for 2′-PP inhibition of Na^+-dependent phosphate uptake was 38 ± 8 nM (Fig. 2). The apparent IC_{50} for 2′-PP inhibition of human intestinal BBMV Na^+-dependent phosphate uptake was similar to that reported for rabbit intestinal brush border membrane vesicles and rat intestinal brush border membrane vesicles [11].

The effect of 2′-PP was specific for the Na^+/phosphate cotransporter and specific for 2′-PP. Na^+-independent phosphate uptake, Na^+-dependent glucose uptake, and Na^+-dependent alanine uptake were not affected by 2′-PP addition to the uptake media (Fig. 2). 4′-PP and 4-PP did not alter Na^+-dependent phosphate uptake into human intestinal brush border membrane vesicles, indicating that the effect of phosphophloretin on the Na^+/phosphate cotransporter was specific for the 2′-isomer (Table 1).

Table 1 indicates that BBMV esterase activity did not contribute to phosphophloretin inhibition of Na^+- dependent phosphate uptake. Although the phosphophloretin derivatives were inhibitors of intestinal BBMV alkaline phosphatase activity, the order of inhibitor potency for phosphophloretin inhibition of alkaline phosphatase was different from the order of inhibitor potency for phosphophloretin inhibition of Na^+- dependent phosphate uptake. Phosphophloretin inhibition of alkaline phosphatase hydrolysis of p-nitrophenyl phosphate followed the sequence: 4-PP > 4′-PP > 2′-PP > phloretin (Table 1). Phosphophloretin inhibition of Na^+-dependent phosphate uptake followed the sequence: 2′-PP ≫ 4′-PP > phloretin > 4-PP (Table 1). The effect of phloretin on human proximal small intestine brush border membrane Na^+-dependent phosphate uptake was similar to the effect of phloretin on Na^+-dependent phosphate uptake into K562 cells and human erythrocytes [24], and into rabbit proximal small intestine brush border membrane vesicles [9].

The effect of external phosphate on 2′-PP inhibition of Na^+-dependent phosphate uptake appeared to be competitive. Increasing external phosphate decreased 2′-PP inhibition of Na^+-dependent phosphate uptake (Fig. 3). Further examination of the results in Fig. 3 by a plot of the slope from Fig. 3 versus the reciprocal of the phosphate concentration and the method of Cornish–Bowden [22,23] indicated mixed inhibition (V_{max} and K_M inhibited).

The effect of 2′-PP on the V_{max} for Na^+-dependent phosphate uptake may be due to the off rate of 2′-PP from the Na^+-loaded cotransporter. If the 2′-PP off rate is much slower than the 3 seconds used for measurements of Na^+-dependent phosphate uptake, the cotransporter: Na^+:2′-PP complex would effectively be a dead-end complex. The resultant removal of a significant percentage of the cotransporter as a dead-end complex would result in a decrease in the apparent transport velocity and decreased V_{max}.

Acknowledgment

The authors gratefully acknowledge the expertise of Edward Ezell, Sealy Center for Structural Biology, UTMB, Galveston, TX, in performing the NMR experiments and in analyzing the results of the NMR experiments.

References

[1] M. Logham-Adham, Phosphate binders for the control of phosphate retention in chronic renal failure, Pedriatr. Nephrol. 13 (1999) 701–708.
[2] E. Slatopolsky, A. Brown, A. Dusso, Pathogenesis of secondary hyperparathyroidism, Kid. Int. 73 (1999) S14–S19.
[3] F. Llach, M. Yudd, Pathogenic, clinical, and therapeutic aspects of secondary hyperparathyroidism in chronic renal failure, Am. J. Kid. Dis. 32 (1998) S3–S12.
[4] F. Llach, J.W. Coburn, in: R.G. Narins (Ed.), Maxwell and Kleeman's Clinical Disorders of Fluid and Electrolyte Metabolism, fifth ed., McGraw-Hill, NY, 1994, Chapter 41.
[5] B.E. Peerce, Identification of the intestinal Na^+-phosphate coransporter, Am. J. Physiol. 256 (1989) G645–G652.
[6] B.E. Peerce, M. Cedilote, S. Seifert, R. Levine, C. Kiesling, R.D. Clarke, Reconstitution of intestinal Na^+-phosphate cotransport, Am. J. Physiol. 264 (1993) G609–G616.
[7] H. Hilfiker, O. Hattenhauer, M. Traebert, I. Forster, H. Murer, J. Biber, Characterization of a murine type II sodium–phosphate cotransporter in mammalian small intestine, Proc. Natl. Acad. Sci. USA 95 (1998) 14564–14569.
[8] H. Xu, L. Bai, J.F. Collins, F.K. Ghishan, Molecular cloning, functional characterization, and chromosomal location of a human, small intestinal sodium–phosphate (Na^+–P_i) transporter, Genomics 62 (1999) 281–284.
[9] B.E. Peerce, R.D. Clarke, A phosphorylated phloretin derivative. Synthesis and effect on intestinal Na^+-dependent phosphate absorption, Am. J. Physiol. 283 (2002) G848–G855.
[10] W. Bernier, R. Kinne, H. Murer, Phosphate transport in brush border membrane vesicles isolated from rat small intestine, Biochem. J. 160 (1976) 467–474.
[11] B.E. Peerce, Examination of the substrate stoichiometry of the intestinal Na^+/phosphate cotransporter, J. Membr. Biol. 110 (1989) 189–197.
[12] L. Chang, B. Sacktor, Sodium gradient-dependent phosphate transport in renal brush border membrane vesicles, J. Biol. Chem. 256 (1981) 1556–1564.
[13] G. Danisi, H. Murer, R.W. Straub, Effect of pH on phosphate transport into intestinal brush border membrane vesicles, Am. J. Physiol. 246 (1984) G180–G186.
[14] S.P. Shirazi-Beechey, J.-P. Gorvel, R.B. Beechey, Phosphate transport in intestinal brush-border membrane, J. Bioenerg. Biomembr. 20 (1988) 273–288.
[15] D.B.N. Lee, M.W. Walling, D.B. Corry, Phosphate transport across rat jejunum: influence of sodium, pH and 1,25-dihydroxyvitamin D_3, Am. J. Physiol. 251 (1986) G90–G95.
[16] A. Dahlquist, Method for assay of intestinal disaccharidases, Anal. Biochem. 7 (1964) 18–25.
[17] S.D. Hanna, A.K. Mircheff, E.M. Wright, Alkaline phosphatase of basal lateral and brush border plasma membranes from intestinal epithelium, J. Supramolec. Struct. 11 (1979) 451–466.
[18] B.S. Furniss, A.J. Hannaford, V. Rogers, P.W.G. Smith, A.R. Tatchell, in: Vogel's Textbook of Practical Organic Chemistry, fourth ed., Wiley, New York, 1978, pp. 782–783.
[19] T. Obata, T. Mukaiyama, A new synthesis of monoalkyl phosphates, J. Org. Chem. 32 (1967) 1063–1065.
[20] K.G. Krebs, D. Heusser, H. Wimmer, in: E. Stahl (Eds.), Thin Layer Chromatography, second ed., Springer, New York, p. 882.
[21] K.G. Krebs, D. Heusser, H. Wimmer, in: E. Stahl (Eds.), Thin Layer Chromatography, second ed., Springer, New York, pp. 886–887.
[22] A.J. Cornish-Bowden, A simple graphical method for determining the inhibition constants of mixed, uncompetitive, and noncompetitive inhibitors, Biochem. J. 137 (1974) 143.
[23] M. Dixon, E.C. Webb, Enzymes, third ed., Academic Press, New York, 1979, Chapter 8.
[24] R.T. Timmer, R.B. Gunn, The molecular basis for Na^+-dependent phosphate transport in human erthryocytes and K562 cells, J. Gen. Physiol. 116 (2000) 363–378.

IX B

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14564–14569, November 1998
Physiology

Characterization of a murine type II sodium-phosphate cotransporter expressed in mammalian small intestine

HELENE HILFIKER, OLAF HATTENHAUER, MARTIN TRAEBERT, IAN FORSTER, HEINI MURER, AND JÜRG BIBER*

Institut of Physiology, University Zürich, CH-8057 Zurich, Switzerland

Communicated by Robert W. Berliner, Yale University School of Medicine, New Haven, CT, September 29, 1998 (received for review July 8, 1998)

ABSTRACT **An isoform of the mammalian renal type II Na/P_i-cotransporter is described. Homology of this isoform to described mammalian and nonmammalian type II cotransporters is between 57 and 75%. Based on major diversities at the C terminus, the new isoform is designed as type IIb Na/P_i-cotransporter. Na/P_i-cotransport mediated by the type IIb cotransporter was studied in oocytes of *Xenopus laevis*. The results indicate that type IIb Na/P_i-cotransport is electrogenic and in contrast to the renal type II isoform of opposite pH dependence. Expression of type IIb mRNA was detected in various tissues, including small intestine. The type IIb protein was detected as a 108-kDa protein by Western blots using isolated small intestinal brush border membranes and by immunohistochemistry was localized at the luminal membrane of mouse enterocytes. Expression of the type IIb protein in the brush borders of enterocytes and transport characteristics suggest that the described type IIb Na/P_i-cotransporter represents a candidate for small intestinal apical Na/P_i-cotransport.**

The kidney and the small intestine are important (externally oriented) control sites to maintain and balance the extracellular concentration of inorganic phosphate (P_i). In the kidney, reabsorption of filtered P_i occurs in the proximal tubule via apically located Na/P_i-cotransporters. Two dissimilar Na/P_i-cotransporters, named type I and type II, have been identified and have been shown to be expressed in the apical membrane of proximal tubular cells (1). As demonstrated recently by targeted inactivation, the type II Na/P_i-cotransporter represents the major pathway by which P_i is reabsorbed (1, 2). With the exception of osteoclasts (3), expression of the type II cotransporter has not yet been described other than in proximal tubules.

In addition to the well characterized renal handling of P_i, an understanding of whole body P_i-homeostasis necessitates elucidating the entry step of P_i in the small intestine (apical Na/P_i-cotransport). However, until now, the molecular identity of a mammalian small intestinal apical Na/P_i-cotransporter has not been described. Although expression of type III Na/P_i-cotransporter mRNA (retroviral receptors Glvr-1 and Ram-1; ref. 4) has been reported in small intestine, as yet, there is no evidence that these Na/P_i-cotransporters are expressed in the apical membrane. Rather, it seems that type III cotransporters are expressed ubiquitously in epithelial and nonepithelial cells.

Based on an expressed sequence tag (EST) clone derived from a cDNA library of murine embryonic cells, we have obtained a functional full length clone coding for a mammalian isoform of the renal type II Na/P_i-cotransporter, which was named type IIb Na/P_i-cotransporter. Expression of type IIb mRNA was found in a variety of tissues, including small intestinal mucosa. By immunohistochemistry, expression of the type IIb protein was localized at the brush border membrane of enterocytes. Transport characteristics of type IIb-mediated Na/P_i-cotransport were similar to the ones described for small intestinal Na/P_i-cotransport (5, 6). Our data suggest that the described type IIb Na/P_i-cotransporter may represent the (a) small intestinal apical Na/P_i-cotransporter.

MATERIALS AND METHODS

Sequencing and Rapid Amplification of 5′-cDNA Ends. An EST-clone (Genome Systems, St. Louis; clone AA647858) with an insert of 3.4 kilobases (kb) was sequenced on both strands. Sequence comparison with the mouse renal type II Na/P_i-cotransporter (7, 8) suggested that ≈700 bp were missing at the 5′ end. To obtain the full length cDNA rapid amplification of 5′ cDNA ends was performed as follows: Total RNA (10 μg) from mouse small intestinal mucosa was retrotranscribed with 200 units MMLV-RT (GIBCO/BRL) by using an oligo-dT_8 primer. Extension of the 5′ end of the cDNA was performed by polynucleotide transferase (30 units, GIBCO/BRL) in the presence of 0.4 mM dATP. PCR was performed with a specific antisense primer derived from the EST-sequence and a T_{17} primer containing a *Sal*I adapter. A second round of PCR was performed with a nested antisense and a Sal*I*-specific primer. The final PCR product was digested with *Sal*I and *Sau*3a and was subcloned into pBluescript SK(+) (Stratagene). The same extension products were obtained by two independent rounds of reactions.

Construction of a Full Length cDNA. Total RNA (10 μg) of mouse small intestine was retrotranscribed by using a dT_8 primer. A PCR fragment was amplified by using a sense primer (nucleotides 8–28) and an antisense primer (nucleotides 961–980) and was cloned into the pGEM-T vector (GIBCO/BRL). The fragment corresponding to the 5′ end of the transporter was excised with *Bgl*II and *Not*I. To obtain the missing 3.2 kb of the transporter, the EST-clone was digested with *Bgl*II and *Sal*I. Both parts were ligated into pSPORT1 (GIBCO/BRL), which was digested with *Not*I and *Sal*I.

Reverse Transcription (RT)–PCR Analysis. Total RNA (10 μg) isolated from different tissues was retrotranscribed with 200 units of reverse transcriptase (MMLV, GIBCO/BRL) by using a dT_8 primer. PCR was performed with a sense (nucleotides 620–638) and an antisense (nucleotides 887–906) primer derived from the EST-clone. Amplification of the mouse kidney-specific type II Na/P_i-cotransporter was achieved by a primer pair derived from the mouse NaP_i-7 cDNA (7, 8): sense position, 10–29; antisense position, 195–210.

Abbreviation: EST, expressed sequence tag; kb, kilobase; RT, reverse transcription.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF081499).

*To whom reprint requests should be addressed at: Institute of Physiology, University Zürich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. e-mail: Biber@Physiol.Unizh.CH.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

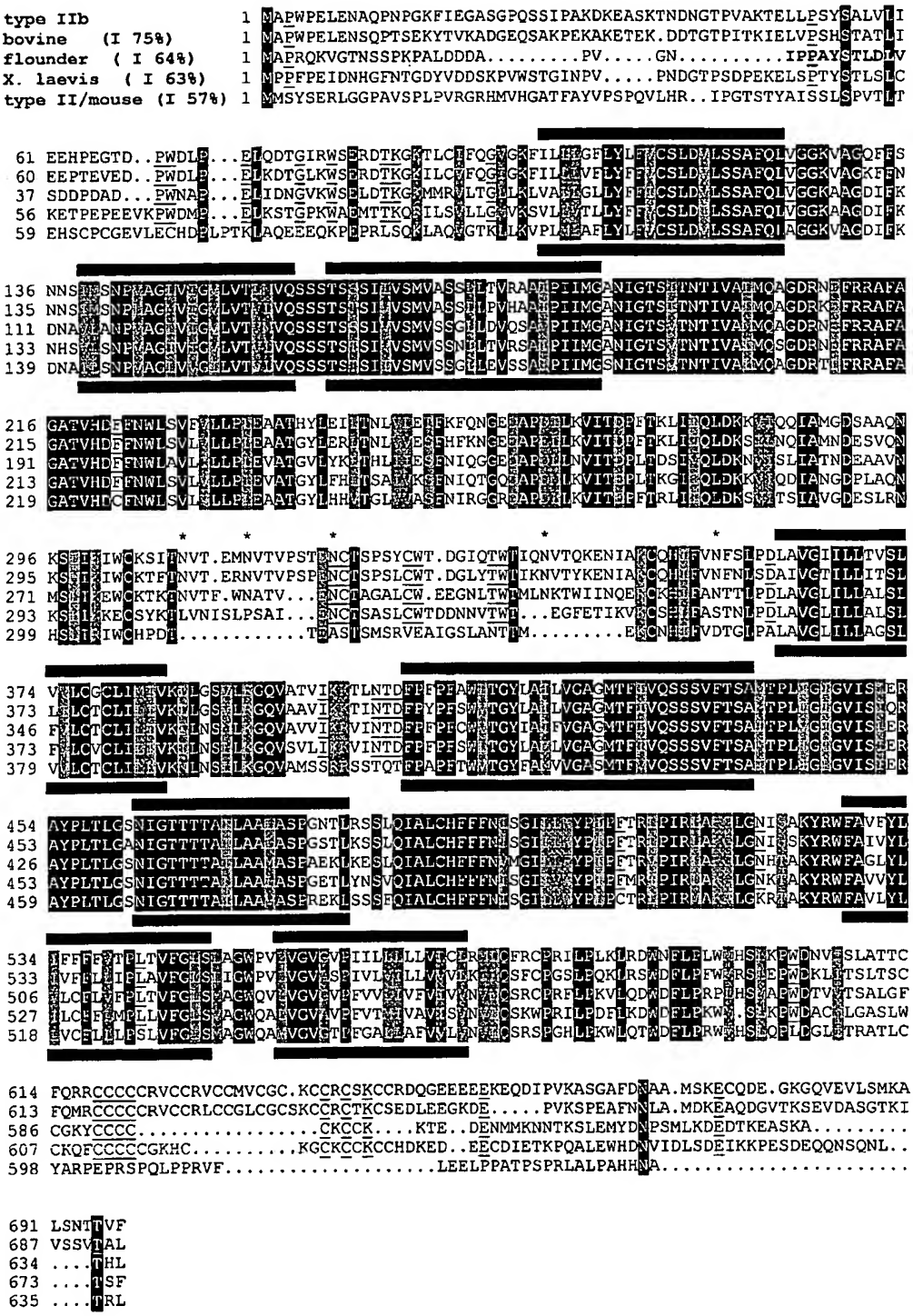


FIG. 1. Amino acid sequence comparisons. Sequences of type II Na/P_i-cotransporters were aligned by pileup (SEQUENCE ANALYSIS software package; Genetics Computer Group, Madison, WI). Shaded boxes indicate consensus residues in all species listed. Additionally, equal residues among the type IIb, bovine, flounder, and *X. laevis* isoforms are underlined. Predicted transmembrane regions are indicated by bars, and potential *N*-glycosylation sites of the proposed extracellular loop are indicated by an asterisk. Numbers given in parentheses (*x*%) indicate the percentage of overall homologies to the type IIb sequence.

Northern Blots. Total RNA from mouse kidney cortex and upper small intestinal mucosa was isolated by the cesium trifluoroacetate/guanidinium thiocyanate method. Poly(A)$^+$ RNA was obtained by oligo-dT cellulose chromatography. Poly(A)$^+$RNA (5 μg) was separated on agarose gels (1.2%) and was transferred onto nylon membranes (BioDyn). Blots were hybridized with the following probes obtained by random priming in the presence of [α^{32}P]dCTP: (*i*) A 5′ end fragment of 900 bp of the type IIb cDNA was obtained by restriction with *Not*I and *Bgl*II, and 9*ii*) a full length probe of the type IIa (NaP_i-7; refs. 7 and 8) cDNA was obtained by restriction with *Not*I and *Sal*I. Equal loadings were confirmed by using probes specific for the ribosomal protein L 28 (9). Hybridization was performed in 6× standard saline citrate (SSC), 5× Denhardt's, 0.5% SDS, and herring sperm DNA (100 μg/ml) at 65°C. Blots were washed sequentially with 2× SSC/0.1% SDS (room

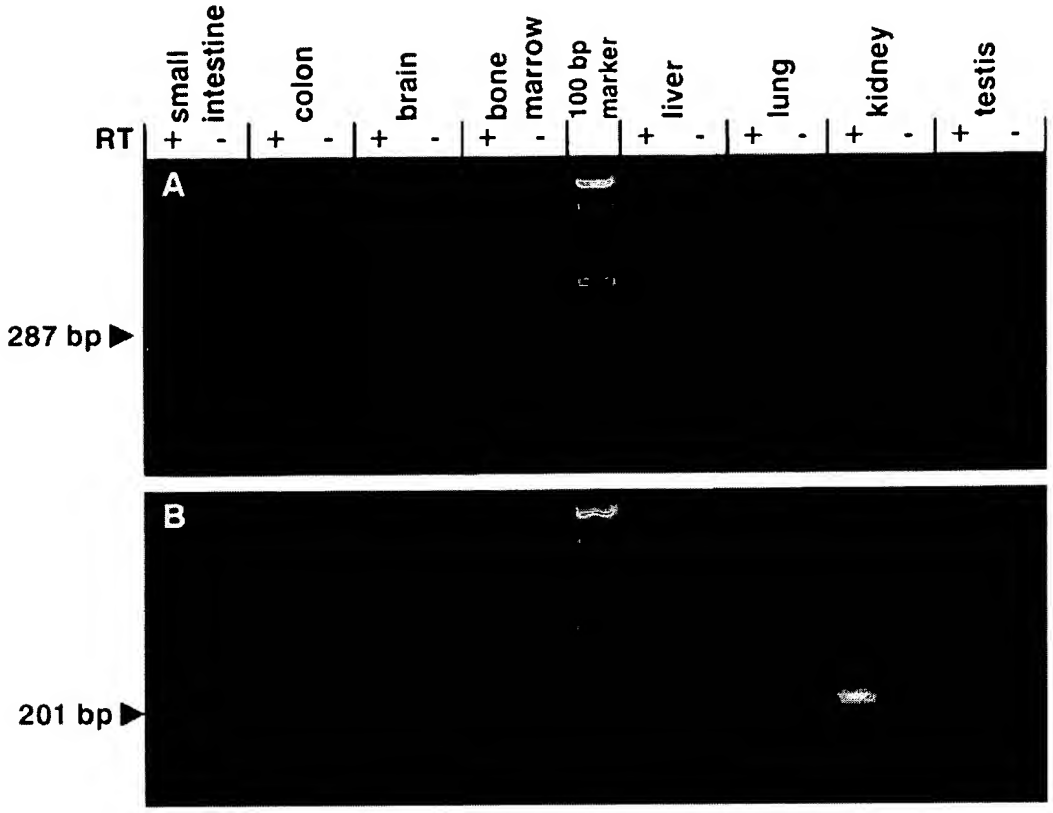


FIG. 2. RT-PCR analysis using primers specific for the type IIb (*A*) or the renal type II (NaP_i-7) (*B*) Na/P_i-cotransporter. All reactions were performed in the presence or absence of reverse transcriptase (RT; +, −). Integrity of the RNA preparations was confirmed by Northern blots using probes specific for β-actin (not shown).

temperature, 10 min), 1× SSC/0.1% SDS (10 min at 40°C), and 0.5× SSC/0.1% SDS (20 min at 55°C). After exposure, results were analyzed by the software package IMAGE QUANT (Molecular Dynamics).

Immunodetections. Rabbit polyclonal antibodies were raised against a synthetic peptide close to the C terminus. Mouse small intestinal brush border membranes were isolated by a Mg^{2+}-precipitation technique (10), and Western blots were performed as described (11). For gel electrophoresis, membranes were denatured in 2% SDS without heating. For immunohistochemistry, mouse duodenum was rinsed with 0.9% NaCl and was fixed by immersion in 3% paraformaldehyde, 0.05% picric acid in a 6:4 mixture of 0.1 M cacodylate buffer (pH 7.4). All other steps were performed as described (11). A swine anti-rabbit IgG conjugated to fluorescein isothiocyanate (Dakopatts, Glostrup, Denmark) was used as a secondary antibody. For peptide protections, the antigenic peptide was added at concentrations of 100 μg/ml.

Transport Assays in Oocytes of *Xenopus laevis*. Isolation and handling of *X. laevis* oocytes has been described elsewhere (12). Oocytes were injected with 5 ng cRNA (in 50nl water). Transport was measured 3 days later, either by isotope flux as described (12, 13, 18) or by electrophysiological means under steady state, voltage clamped conditions (14).

RESULTS

By databank search, an EST cDNA clone (AA647858) prepared from mouse two-cell stage embryos was found that showed 73% homology over a length of 312 bp to the mouse renal type II Na/P_i-cotransporter (7, 8). Preliminary analysis by Northern blotting indicated that a related mRNA species is expressed in small intestinal mucosa (data not shown). Full length sequencing of the EST clone (3.5 kb) suggested that, at the 5′ end, 700 bp were missing. By rapid amplification of 5′ cDNA ends, a full length cDNA (4,039 bp) was obtained (GenBank accession no. AF081499) containing an ORF (positions 45–2,137) coding for a protein of 697 amino acids (Fig. 1).

Amino acid comparisons revealed that the newly identified protein is 57–75% homologous to Na/P_i-cotransporters identified in bovine NBL cells (15), flounder kidney and intestine (16), and intestine and lung of *X. laevis* (17) and to the renal type II Na/P_i-cotransporter (NaP_i-7; refs. 7 and 8) (Fig. 1). Overall homology to type I (1) and type III Na/P_i-cotransporters (4) was ≈20%. As illustrated, highest homologies among the listed Na/P_i-cotransporters are seen in regions that also have been proposed to represent transmembrane regions (1). The most striking difference of the newly identified protein compared with the mouse renal type II Na/P_i-cotransporter is found in the C-terminal region containing clusters of cysteine residues. A similar clustering of cysteine residues is also present in the Na/P_i-cotransporters of bovine cells, flounder kidney/intestine, and *Xenopus* intestine. Therefore, we propose to subdivide type II Na/P_i-cotransporters into a subfamily type IIa (represented by the renal isoforms of mouse, rat, rabbit, opossum kidney cells, and human; ref. 1) and type IIb (represented by the isoforms of bovine, flounder, and *Xenopus* and the one described here).

Expression of type IIb mRNA was analyzed by RT-PCR using total RNA (Fig. 2) and Northern blots using poly(A)$^+$RNA (Fig. 3). By RT-PCR using primers positioned within the ORF, expression of type IIb mRNA was indicated in the mucosa of the upper small intestine, colon, liver, lung, kidney, and testis. As a control, the same RNA samples were subjected to RT-PCR analysis for the renal type IIa cotransporter NaP_i-7. As indicated, a type IIa-related PCR product was found only in RNA isolated from kidney, confirming the

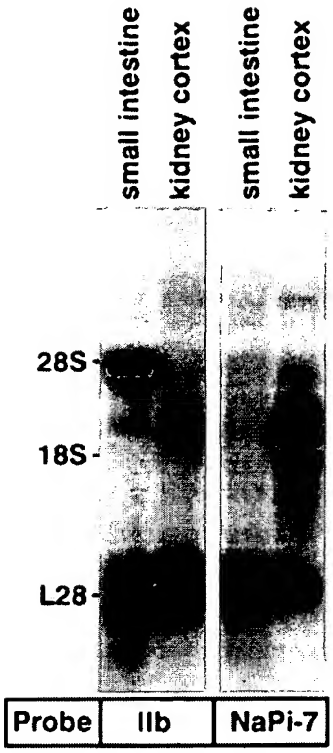


FIG. 3. Northern blot analysis of poly(A)+RNA isolated from mouse upper small intestinal mucosa and kidney cortex. Blots were hybridized with probes derived from a 900-bp 5′ end fragment of type IIb cDNA or from the full length cDNA of the mouse renal type II Na/P_i-cotransporter. Hybridization to the ribosomal protein L28 mRNA was used to confirm equal loadings. In the case of the NaP_i-7 probe, five times less poly(A)+RNA of kidney cortex was loaded.

kidney specific expression of the type IIa cotransporter (1, 18). Northern blots performed with poly(A)$^+$RNA isolated of mouse kidney cortex and small intestine are shown in Fig. 3. By using a 5′-end probe of 900 bp, the major mRNA species detected in small intestinal mRNA was at ≈4 kb. In mRNA of kidney cortex, no such signal was detectable. In addition, in small intestinal mRNA, a faint signal at ≈2.5 kb was evident.

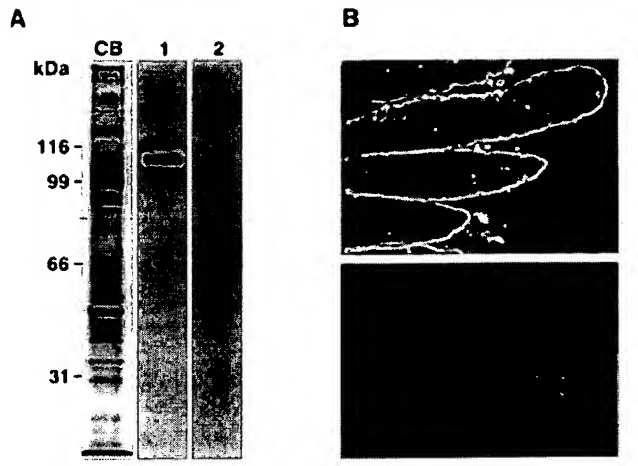


FIG. 4. Immunodetection of the type IIb Na/P_i-cotransporter. (*A*) Western blots of isolated mouse small intestinal brush border membranes (35 μg protein per lane). CB, Coomassie blue staining. Incubation with the first antibody was performed in the absence (lane 1) or presence (lane 2) of the antigenic peptide. (*B*) Immunofluorescence detection of the type IIb cotransporter in the apical membrane of enterocytes. Incubation with the primary antibody was performed in the absence (*Upper*) or presence (*Lower*) of antigenic peptide (100 μg/ml).

By using the same probe, two signals at ≈2.5 kb also were detected with poly(A)$^+$RNA of mouse kidney cortex. To verify a possible crossreaction with the renal type II cotransporter, the same blots were hybridized with probes derived from the NaP_i-7 cDNA (Fig. 3*B*). As illustrated, no signals with small intestinal mRNA were detected by this probe, but, with mouse kidney cortex mRNA, a strong signal (double band at ≈2.5 kb) was observed representing the renal type II Na/P_i-cotransporter. This suggested that the double band seen in kidney mRNA with the type IIb probe represents a crossreaction with the type II (NaP_i-7) cotransporter. Confirmation for such a crossreaction was obtained with poly(A)$^+$RNA isolated from kidney cortex of mice in which the renal type II Na/P_i-cotransporter has been knocked out (19).

Expression of the type IIb protein was analyzed by immunoblotting and immunofluorescence using a polyclonal anti-

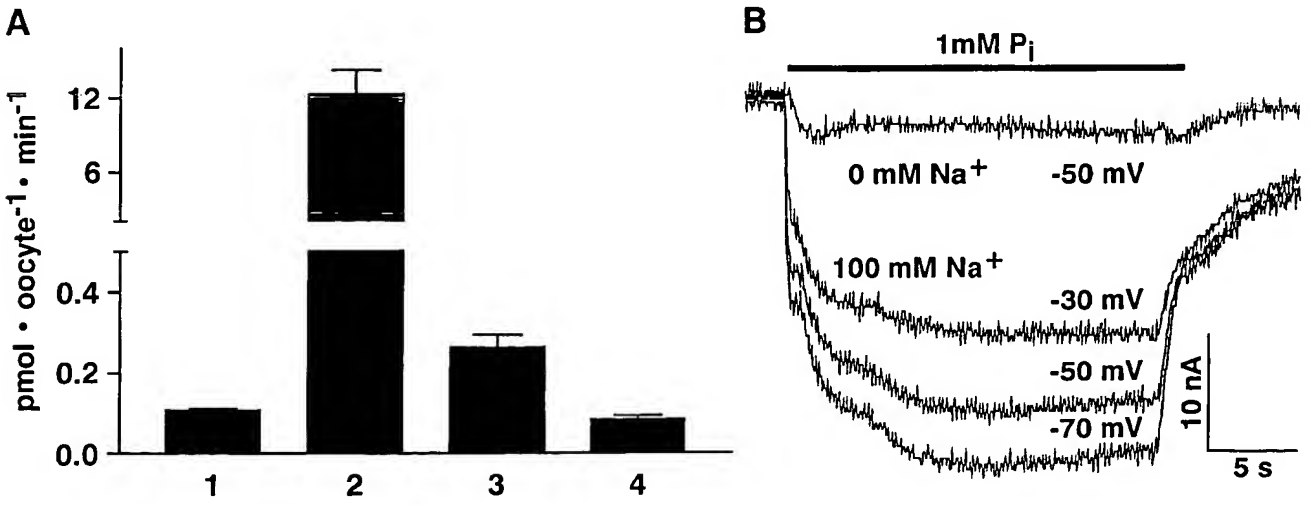


FIG. 5. Characterization of Na/P_i-cotransport in oocytes injected with type IIb cRNA. (*A*) Isotope flux measurements performed at 0.5 mM phosphate (1, 2, 3) or 1 mM sulfate (4) (mean ± SD of 8–10 oocytes; two experiments). Bar 1: P_i-uptake in the presence of NaCl into oocytes injected with water; bars 2 and 3: P_i-uptake into oocytes injected with type IIb cRNA in the presence of NaCl (2) or choline-Cl (3). (*B*) Inwardly directed currents measured under steady state conditions by using the two-electrode voltage clamp. Oocytes were voltage clamped at indicated voltages and were superfused with 1 mM phosphate in the absence or presence of NaCl (see refs. 8 and 14).

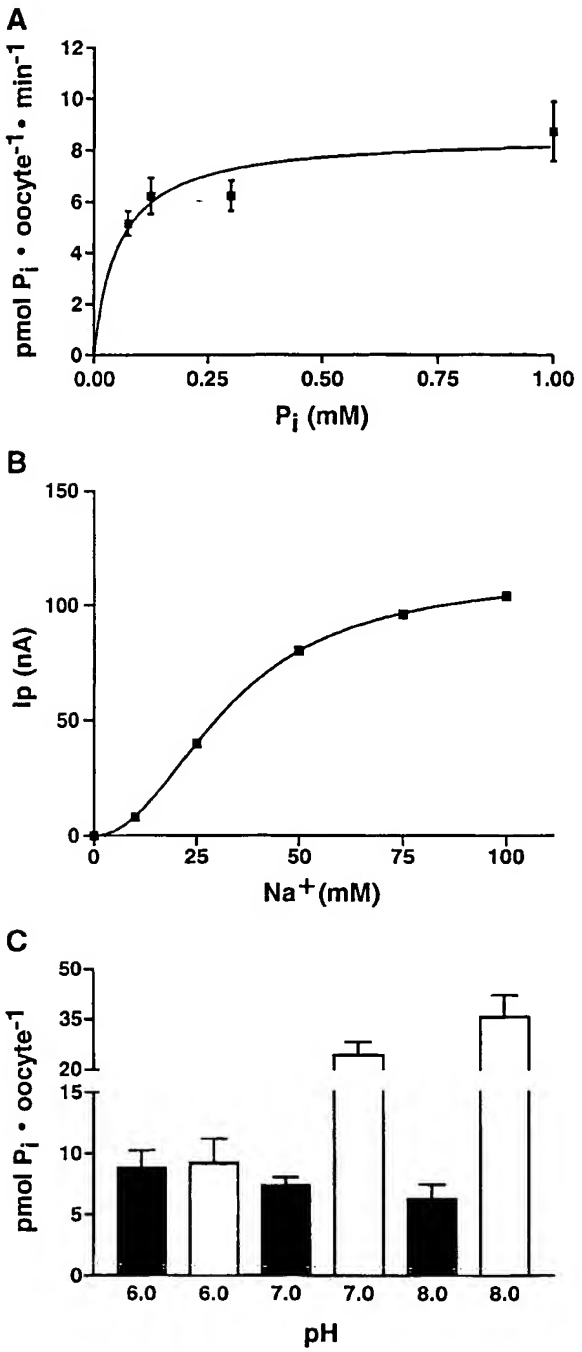


FIG. 6. Characterization of type IIb-mediated Na/P_i-cotransport. (*A*) Isotope flux measurements were performed at different P_i concentrations in the presence of sodium and were corrected by the values obtained from oocytes injected with water. The calculated value for $K_m(P_i)$ was 50 μM. (*B*) Electrophysiological determination of the K_m(Na). The calculated K_m(Na) was 33 mM, and the stoichiometry was >2. (*C*) pH dependence of type IIb (black bars) and the renal type II (NaP_i-7) Na/P_i-cotransport (open bars). Values were corrected with uptake in the presence of choline-Cl, which was not changed by the different pH values. The data represent the mean ± SD of 8–10 oocytes. All experiments have been performed at least twice. The data given in *C* were derived from one oocyte. The same result was obtained with different oocytes from different batches.

body raised against a synthetic C-terminal peptide. On Western blots performed with isolated small intestinal brush border membranes, a reaction with a single band of ≈108 kDa was observed that could be protected completely by inclusion of the antigenic peptide (Fig. 4*A*). Because four potential *N*-glycosylation sites are contained in a region representing an extracellular loop and because, in the case of the renal type II cotransporter, *N*-glycosylation has been demonstrated in this loop (20), the molecular mass of 108 kDa likely represents the glycosylated form of the type IIb protein (unglycosylated M_r 78). In cryostat sections of mouse duodenum, specific reaction was observed at the apical membrane of enterocytes that was prevented by the antigenic peptide (Fig. 4*B*).

Type IIb cRNA was injected into oocytes of *X. laevis*, and transport of phosphate was measured in the presence and absence of sodium by isotope flux (Fig. 5*A*). Compared with oocytes injected with water, oocytes injected with type IIb cRNA exhibited a large expression of P_i-transport, which depended on the presence of sodium and which was not observed after injection of antisense cRNA (data not shown). Furthermore, oocytes injected with type II cRNA did not take up sulfate (SO_4^{2-}), suggesting that the type IIb cotransporter exhibits similar specificity as described for the renal type IIa cotransporter (13, 18). As reported for the renal type II Na/P_i-cotransporter (8, 14), superfusion of oocytes expressing the type IIb cotransporter with phosphate exhibited an inwardly directed current that depended on the presence of sodium and the steady state holding potential (Fig. 5*B*). This indicated that, like the type II cotransporter, Na/P_i-cotransport by the type IIb cotransporter is electrogenic. As observed in isotope flux measurements (Fig. 5*A*), there was also evidence for a small contribution of a Na-independent P_i-transport. Based on the results shown in Fig. 5, transport characteristics of type IIb-induced Na/P_i-cotransport were determined by either isotope flux or by electrophysiological measurements. By both methods an apparent $K_m(P_i)$ of ≈50 μM and a K_m(Na) of ≈30 mM was determined (Fig. 6). Because pH dependence is a hallmark of the renal type II Na/P_i-cotransporter, type IIb-mediated Na/P_i-cotransport was determined at different pH-values (Fig. 6*C*). In contrast to the renal type II isoform, type IIb-associated Na/P_i-cotransport was less dependent of the pH and was slightly higher at more acid pH-values.

DISCUSSION

Intake and extrusion of inorganic phosphate is determined by the intestinal and renal handling of phosphate. In recent years, some renal phosphate transporters and their roles in the regulation of the renal handling of P_i have been described (1, 2, 4, 7, 8, 18). However, small intestinal phosphate cotransporters expressed in the mammalian small intestine are characterized far less.

With respect to the renal handling of P_i, three dissimilar types of sodium-dependent phosphate cotransporters expressed in the plasma membrane have been described so far: a type I, a type II, and a type III (1). It could be shown that the type II Na/P_i-cotransporter plays a major role in proximal tubular reabsorption of P_i and that proximal tubular capacity to reabsorb P_i to a large part depends on the net abundance of type II cotransporters in the apical membrane of proximal tubules (1, 2). The role of type I and III cotransporters is less clear.

So far, expression of type II mRNA has been described in kidney cortex only. Expression of type I mRNA also is found in liver and brain, and expression of type III mRNA seems to occur in almost every tissue (1, 4), notably, also, in the intestine. The nonspecific tissue expression pattern of type III cotransporters rules out the possibility of type III Na/P_i-cotransporters being candidates for small intestinal apical Na/P_i-cotransporters.

Derived from an EST clone from an embryonic mouse cDNA library, we obtained the full length cDNA coding for a type II Na/P_i-cotransporter. The deduced protein showed high

homology to described type II Na/P_i-cotransporters. Highest homologies were found in regions that most likely represent transmembrane segments, and the most striking differences between the identified cotransporter, the bovine, flounder, and *X. Laevis* isoforms, and the renal type Na/P_i-cotransporters are represented by cysteine clusters at the C-terminal ends. Therefore, we propose the name type IIb Na/P_i-cotransporter for the identified mammalian isoform and propose to extend this nomenclature also for the bovine (15), the flounder, (16) and the *X. laevis* (17) isoforms.

Injection of type IIb cRNA into oocytes of *X. laevis* resulted in expression of Na-dependent phosphate transport with characteristics similar to that observed for Na/P_i-cotransport mediated by the renal type II Na/P_i-cotransporter (8, 14, 18). However, the most striking difference of type IIb-mediated Na/P_i-cotransport was its pH dependence.

So far, proteins involved in mammalian small intestinal Na/P_i-cotransport have not been described. In nonruminants, highest rates of P_i-reabsorption are observed in the upper small intestine (5). Na/P_i-cotransport in mouse small intestine is highest at a more acidic pH and exhibits a K_m value for P_i of $\approx 50\ \mu M$ (5, 6).The functional characteristics observed for type IIb-mediated Na/P_i-cotransport are in agreement with these data and support the notion that the type IIb cotransporter may represent a candidate for a small intestinal Na/P_i-cotransporter. This is supported further by the observation that both type IIb mRNA and protein are expressed in mouse small intestinal mucosa and, notably, that the type IIb protein is localized at the brush border membrane of the enterocytes.

In summary, we have identified a mammalian Na/P_i-cotransporter with high homology to described type II Na/P_i-cotransporters. Expression of the mRNA and the protein of such a transport protein was demonstrated in the mammalian small intestine. Kinetic properties and pH dependence of type IIb-associated Na/P_i-cotransport favor this protein as a candidate for a Na/P_i-cotransporter involved in intestinal P_i-reabsorption. Apart from the small intestine, expression of type IIb mRNA also was recognized in other tissues, such as lung, colon, liver, kidney, and testis. The physiological role of Na/P_i-cotransport mediated by the type IIb isoform in small intestine as well as in the other tissues remains to be determined.

We thank Drs. R. Wenger and M. Gassmann for providing RNA samples of different mouse tissues. The art work of C. Gasser is gratefully acknowledged. Financial support was provided by the grants of the Swiss National Fonds to J.B. and H.M. and from the University of Zürich (Stiftung für wissenschaftliche Forschung).

1. Murer, H. & Biber, J. (1997) *Eur. J. Physiol.* **433,** 379–389.
2. Levi, M., Kempson, S. A., Lötscher, M., Biber, J. & Murer, H. (1996) *J. Membr. Biol.* **154,** 1–9.
3. Gupta, A., Guo, X. L., Alvarez, U. M. & Hruska, K. A. (1997) *J. Clin. Invest.* **100,** 539–549.
4. Kavanaugh, M. P. & Kabat, D. (1996) *Kidney Int.* **49,** 959–963.
5. Cross, H. S., Debiec, H. & Peterlik, M. (1990) *Miner. Electrolyte Metab.* **16,** 115–124.
6. Nakagawa, N. & Ghishan, F. K. (1993) *Proc. Soc. Exp. Biol. Med.* **203,** 328–335.
7. Collins, J. F. & Ghishan F. K. (1994) *FASEB J.* **8,** 862–868.
8. Hartmann, C., Wagner, C. A., Busch, A. E., Markovich, D., Biber, J., Lang, F. & Murer, H. (1995) *Pflügers Arch.* **430,** 830–836.
9. Burke, P. S., Lium, E. & Lin, C. S. (1994) *Gene* **142,** 315–316.
10. Stieger, B. & Murer, H. (1983) *Eur. J. Biochem.* **135,** 95–101.
11. Custer, M., Lötscher, M., Biber, J., Murer, H. & Kaissling, B. (1994) *Am. J. Physiol.* **266,** F767–F774.
12. Werner., A., Biber, J., Forgo, J., Palacin, M. & Murer, H. (1990) *J. Biol. Chem.* **265,** 12331–12336.
13. Markovich, D., Forgo, J., Stange, G., Biber, J. & Murer, H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8073–8077.
14. Forster, I. C., Wagner, C. A., Busch, A. E., Lang, F., Biber, J., Hernando, N., Murer, H. & Werner, A. (1997) *J. Membr. Biol.* **160,** 9–25.
15. Helps, C., Murer, H. & McGivan, J. (1995) *Eur. J. Biochem.* **228,** 927–930.
16. Kohl, B., Herter, P., Hulseweh, B., Elger, M., Hentschel, H., Kinne, R. K. & Werner, A. (1996) *Am. J. Physiol.* **270,** F937–F944.
17. Ishizuya-Oka, A., Stolow, M. A., Ueda, S. & Shi Y. B. (1997) *Dev. Genetics* **20,** 53–66.
18. Magagnin, S., Werner, A., Markovich, D., Sorribas, V., Biber, J. & Murer, H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5979–5983.
19. Beck, L., Karaplis, A. C., Amizuka, N., Hewson, A. S., Ozawa, H. & Tenenhouse, H. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5327–5377.
20. Hayes, G., Busch, A., Lötscher, M., Waldegger, S., Lang, F., Verrey, F., Biber, J. & Murer, H. (1994) *J. Biol. Chem.* **269,** 24143–24149.

SHORT COMMUNICATION

Molecular Cloning, Functional Characterization, Tissue Distribution, and Chromosomal Localization of a Human, Small Intestinal Sodium–Phosphate (Na^+–P_i) Transporter (SLC34A2)

Hua Xu, Liqun Bai, James F. Collins, and Fayez K. Ghishan[1]

Department of Pediatrics and Department of Physiology, Steele Memorial Children's Research Center, University of Arizona Health Sciences Center, Tucson, Arizona 85724

Received June 4, 1999; accepted August 26, 1999

Phosphate plays a crucial role in cellular metabolism, and its homeostatic regulation in intestinal and renal epithelia is critical. Apically expressed sodium–phosphate (Na^+–P_i) transporters play a critical role in this regulation. We have isolated a cDNA (HGMW-approved symbol SLC34A2) encoding a novel human small intestinal Na^+–P_i transporter. The cDNA is shown to be 4135 bp in length with an open reading frame that predicts a 689-amino-acid polypeptide. The putative protein has 76% homology to mouse intestinal type II Na^+–P_i transporter (Na/Pi-IIb) and lower homologies with renal type II Na^+–P_i transporters. Northern blots showed a singular transcript of 5.0 kb in human lung, small intestine, and kidney. Computer analysis suggests a protein with 11 transmembrane domains and several potential posttranslational modification sites. Functional characterization in *Xenopus laevis* oocytes showed that this cDNA encodes a functional Na^+–P_i transporter. Furthermore, the gene encoding this cDNA was mapped to human chromosome 4p15.1–p15.3 by the FISH method. © 1999 Academic Press

Phosphate (P_i) plays a major role in growth, development, bone formation, and cellular metabolism. The kidney and small intestine are important regulatory sites that maintain extracellular P_i concentrations. Sodium-coupled phosphate transport is the major form of P_i absorption in both kidney and intestine. Phosphate uptake by renal and intestinal brush-border membrane vesicles has been studied previously in human (2), rat (6), rabbit (3), and mouse (15). The molecular basis of P_i uptake in kidney has been identified (sodium–phosphate [Na^+–P_i] transporters types I and II) and well characterized. Type II Na^+–P_i transporter is the major transport pathway of P_i reabsorption in kidney (5, 12). However, little is known about P_i absorption in the intestine. To date, only one mammalian, intestinal Na^+–P_i transporter has been identified from mouse (10).

To further define the role of the small intestinal Na^+–P_i transporter (NPT) in body P_i homeostasis, we have isolated a novel cDNA[2] from a human small intestinal cDNA library. Initially, we designed PCR primers (homologous to mouse intestinal Na/Pi-IIb cDNA; forward primer at 686–705 bp and reverse primer at 1423–1442 bp) for RT-PCR experiments. PCR products of the predicted size (760 bp) were obtained from human small intestinal cDNA (mRNA

[2] The HGMW-approved symbol for the gene described in this paper is SLC34A2.

```
MAPWPELGDAQPNPDKYLEGAAGQQPTAPDKSKETNKNNTEAPVTKIELL  50
PSYSTATLIDEPTEVDDPWNLPTLQDSGIKWSERDTKGKILCFFQGIGRL 100
          TM1                                   TM2
ILLLGFLYFFVCSLDILSSAFQLVGGKMAGQFFSNSSIMSNPLLGLVIGV 150
                                          TM3
LVTVLVQSSSTSTSIVVSMVSSSLLTVRAAIPIIMGANIGTSITNTIVAL 200
                              TM4
MQVGDRSEFRRAFAGATVHDFFNWLSLLVLLPVEVATHYLEIITQLIVES 250
FHFKNGEDAPDLLKVITKPFTKLIVQLDKKVISQIAMNDEKAKNKSLVKI 300
WCKTFTNKTQINVTVPSTANCTSPSLCWTDGIQNWTMKNVTYKENIAKCQ 350
                    TM5
HIFVNFHLPDLAVGTILLILSLLVLCGCLIMIVKILGSVLKGQVATVIKK 400
               TM6                          TM7
TINTDFPFPFAWLTGYLAILVGAGMTFIVQSSSVFTSALTPLIGIGVITI 450
               TM8                          TM9
ERAYPLTLGSNIGTTTTAILAALASPGNALRSSLQIALCHFFFNISGILL 500
                                        TM10
WYPIPFTRLPIRMAKGLGNISAKYRWFAVFYLIIFFFLIPLTVFGLSLAG 550
          TM11
WRVLVGVGVPVVFIIILVLCLRLLQSRCPRVLPKKLQNWNFLPLWMRSLK 600
PWDAVVSKFTGCFQMRCCCCCRVCCRACCLLCGCPKCCRCSKCCEDLEEA 650
QEGQDVPVKAPETFDNITISREAQGEVPASDSKTECTAL             689
```

FIG. 1. Predicted amino acid sequence of human small intestinal NPT. Underlined amino acid sequences represent potential transmembrane (TM) domain regions, which are numbered sequentially. Boldface amino acids represent putative N-glycosylation sites. Nucleotide sequence can be found in GenBank under Accession No. AF146796.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF146796.

[1] To whom correspondence should be addressed at Department of Pediatrics, Steele Memorial Children's Research Center, 1501 N. Campbell Avenue, Tucson, AZ 85724. Telephone: (520) 626-5170. Fax: (520) 626-4141. E-mail: fghishan@peds.arizona.edu.

Genomics 62, 281–284 (1999)
Article ID geno.1999.6009
0888-7543/99 $30.00

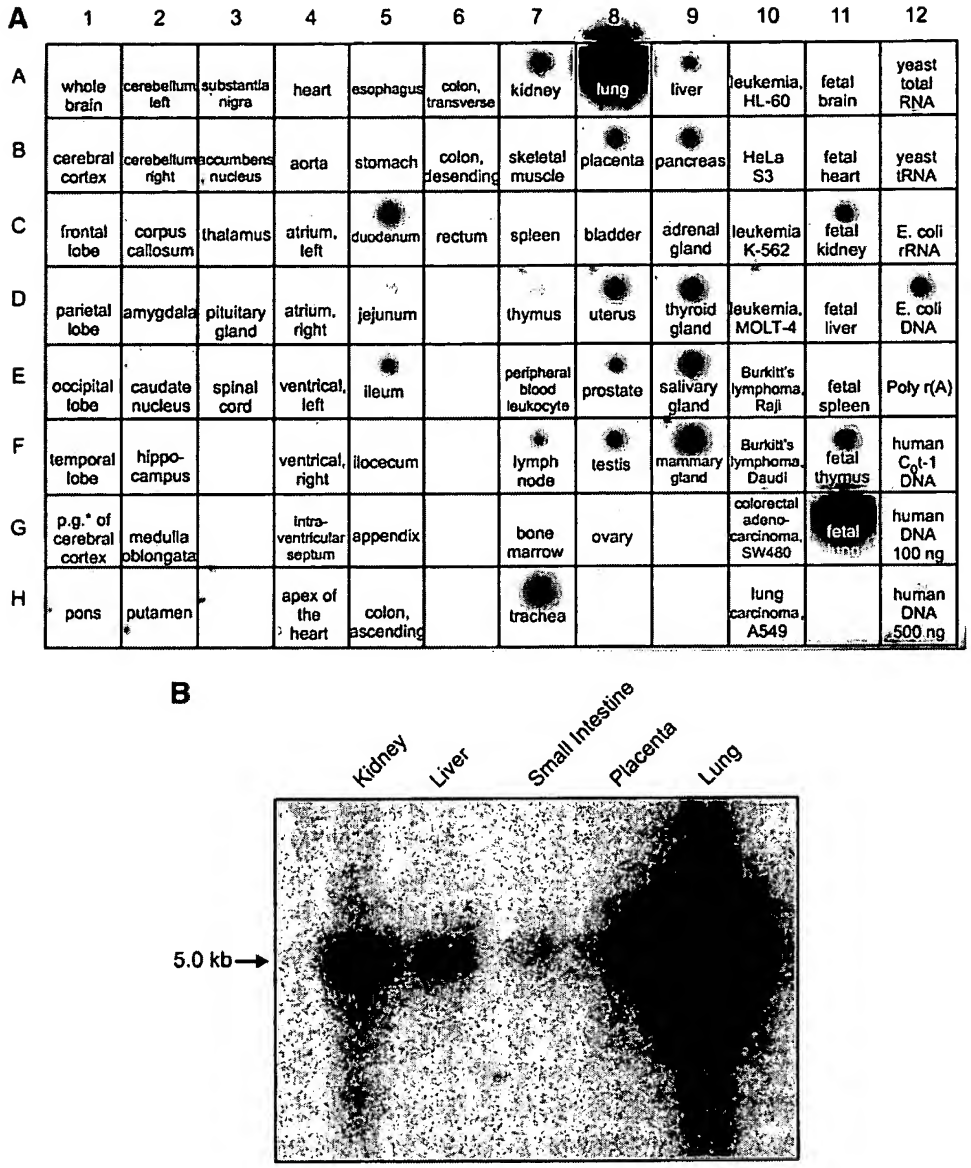


FIG. 2. mRNA expression of human intestinal NPT. mRNA from 76 different human tissues was loaded on the nylon membrane [mRNA levels were normalized versus eight different gene markers (Clontech manual), so this blot is quantitative] (A). The blot was hybridized with cDNA-specific probes under high-stringency conditions and exposed to film. Also shown is a Northern blot of human small intestinal NPT (B). The blot was hybridized with cDNA-specific probes under high-stringency conditions and exposed to film. A 5-kb transcript was detected in several tissues. The blot is not quantitative, as mRNA loading was not normalized.

from Clontech, Palo Alto, CA). Sequencing of the amplified fragment revealed 82% nucleotide sequence identity with the mouse intestinal Na/Pi-IIb. This suggested that this cDNA likely represents a type II Na^+–P_i transporter in human small intestine.

Using the PCR product to generate radioactive probes, we isolated a cDNA clone from an enriched, human small intestinal cDNA library (Edge Biosystems, Gaithersburg, MD), utilizing high-stringency screening conditions. This cDNA was sequenced on both strands. Sequence data indicated that this human intestinal cDNA has 4135 bp and encodes a putative protein of 689 amino acids (open reading frame 36 to 2102 bp) (Fig. 1). Hydropathy analysis (Omiga 1.1.3 software, Oxford Molecular Ltd., Oxford, England) predicts 11 transmembrane domains. The putative protein has many potential posttranslational modification sites. We compared our cDNA with other identified Na^+–P_i transporter cDNAs, which showed over 60% nucleotide sequence identity and over 75% amino acid sequence identity with bovine renal type II Na^+–P_i transporter (7), mouse intestinal type II Na^+–P_i transporter (10), and an unpublished human NPT (GenBank Accession No. AF111856). Recently, another type I Na^+–P_i transporter cDNA was isolated from a human intestinal cDNA library (16); however, the sequence similarity with our cDNA is very low (<20%). Overall, these findings suggest that our newly identified human cDNA clone encodes a protein that belongs to the intestinal type II Na^+–P_i transporter gene family.

mRNA expression of this cDNA clone was analyzed with a human 76-tissue mRNA blot (Human Multiple Tissue Expression Array, Clontech). Hybridization utilizing the 760-bp PCR fragment as a template to generate radiolabeled probes showed that this gene is expressed in many tissues (Fig. 2A). Highest expression was seen in lung, small intestine, and kidney. Furthermore, Northern blot analysis [human 11-tissue mRNA blot (Clontech)] showed a single transcript at approximately 5.0 kb in human lung, small intestine, kidney, liver, and placenta (Fig. 2B). Interestingly, previous studies showed that human renal NPT mRNA transcripts were detected at 2.0 kb [type I Na^+–P_i transporter (14)] and at 2.7 kb [type II Na^+–P_i transporter (12)]. This suggests that this 5-kb transcript detected in human kidney likely represents an unidentified type II Na^+–P_i transporter isoform. Also, since this gene is highly expressed in adult and fetal lung, it seems probable that this newly identified human Na^+–P_i transporter has an important physiological function in lung (possibly involved in the production of surfactant by the alveoli).

To characterize the function of the protein encoded by this cDNA, we produced cRNA, injected it into *Xenopus laevis* oocytes, and measured radiolabeled P_i influx in the presence or in the absence of Na^+ (1, 5). Compared with uninjected oocytes, the Na^+–P_i transporter cRNA injected oocytes exhibited an approximate 55-fold increase of P_i transport ($P < 0.0001$) (Fig 3), which suggested that this cDNA does indeed encode a functional Na^+–P_i transporter.

To identify the locus of our Na^+–P_i transporter gene, a PCR-amplified 760-bp cDNA fragment was used to localize the gene position by performing the FISH mapping technique in lymphocytes isolated from human blood (SeeDNA Biotech Inc., Windsor, Ontario, Canada). The 760-bp human intestinal NPT cDNA probe

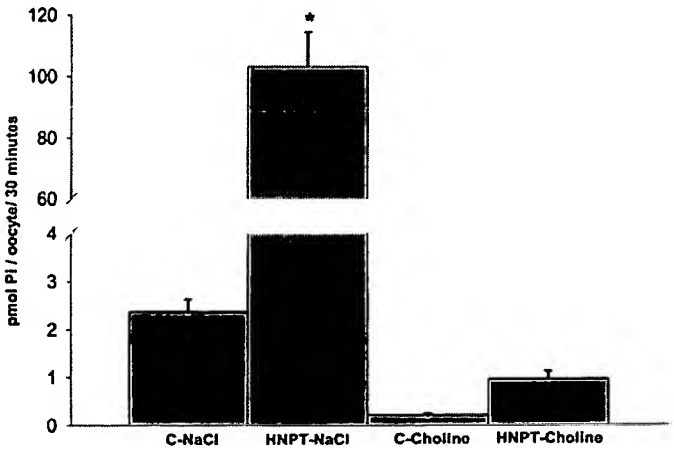


FIG. 3. Characterization of human intestinal type II NPT in oocytes injected with cRNA. Isotope (P_i) influx measurements were performed in the presence of NaCl or choline chloride. An approximate 55-fold increase of P_i uptake was seen in cRNA-injected oocytes. C, uninjected oocytes. HNPT, cRNA-injected oocytes. (Mean ± SEM of 5 oocytes per group; $n = 4$, $^*P < 0.0001$ for HNPT versus all other groups).

FIG. 4. Chromosomal localization of the human small intestinal NPT gene to chromosome 4p15.1–p15.3. (**Left**, arrow) FISH signals on human chromosome hybridized with the 760-bp probe. (**Right**) The same mitotic chromosome stained with DAPI to identify human chromosome 4.

was biotinylated with dATP (8), and FISH detection was performed by the procedure described by Heng and Tsui (9). FISH signals and the DAPI-banding pattern were recorded separately, and the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals with DAPI-banded chromosomes. Results showed that hybridization signal was detected only on chromosome 4p15.1–p15.3 (Fig. 4). Previous investigations showed that human renal Na^+–P_i transporters are found on chromosomes 5 and 6 [5q35 (11, 13), 6p21.3 (unpublished results; GenBank Accession Nos. U90544 and U90545), and 6p21.1–p23 (4, 11)]. Also, type I Na^+–P_i transporter cDNA from human intestinal mucosa was localized to 6p21.3–p23 (16). These data indicate that Na^+–P_i transporter genes are widely distributed in the genome.

In summary, we isolated a 4135-bp type II Na^+–P_i transporter cDNA from a human small intestinal cDNA library. This cDNA encodes a 689-amino-acid protein, which shares high sequence homology with intestinal and renal type II Na^+–P_i cDNAs from various mammalian species. This gene encodes a Na^+–P_i transporter that is highly expressed in human lung, small intestine, and kidney and is localized to chromosome 4p15.1–p15.3.

ACKNOWLEDGMENT

This investigation was funded by NIH Grant R01-DK33209-17.

Note added in proof. The unpublished human Na^+–P_i transporter cDNA sequence mentioned in the text (Genbank Accession No. AF111856) has now been published (Field, J. A., Zhang, L., Brun, K. A., Brooks, D. P., and Edwards, R. M. (1999). Cloning and functional characterization of a sodium-dependent phosphate transporter expressed in human lung and small intestine. *Biochem. Biophys. Res. Commun.* **258**: 578–582).

REFERENCES

1. Bai, L., and Pajor, A. M. (1997). Expression cloning of NaDC-2, an intestinal Na(+) or Li(+)-dependent dicarboxylate transporter. *Am. J. Physiol.* **273:** G267–G274.
2. Borowitz, S. M., and Ghishan, F. K. (1989). Phosphate transport in human jejunal brush-border membrane vesicles. *Gastroenterology* **96:** 4–10.
3. Borowitz, S. M., and Granrud, G. S. (1992). Ontogeny of intestinal phosphate absorption in rabbits. *Am. J. Physiol.* **262:** G847.
4. Chong, S. S., Kristjansson, K., Zoghbi, H. Y., and Hughes, M. R. (1993). Molecular cloning of the cDNA encoding a human renal sodium phosphate transporter protein and its assignment to chromosome 6p21.2–p23. *Genomics* **18:** 355–359.
5. Collins, J. F., and Ghishan, F. K. (1994). Molecular cloning, functional expression, tissue distribution, and in situ hybridization of the renal sodium phosphate (Na/P_i) transporter in the control and hypophosphatemic mouse. *FASEB J.* **8:** 862–868.
6. Ghishan, F. K., Arab, N., and Shibata, H. (1990). Intestinal phosphate transport in spontaneously hypertensive rats and genetically matched controls. *Gastroenterology* **99:** 106–112.
7. Helps, C., Murer, H., and McGiven, J. (1995). Cloning, sequence analysis and expression of the cDNA encoding a sodium-dependent phosphate transporter from the bovine renal epithelial cell line NBL-1. *Eur. J. Biochem.* **228:** 927–930.
8. Heng, H. H. Q., Squire, J., and Tsui, L. C. (1992). High resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Proc. Natl. Acad. Sci. USA* **89:** 9509–9513.
9. Heng, H. H. Q., and Tsui, L. C. (1993). Modes of DAPI banding and simultaneous in situ hybridization. *Chromosoma* **102:** 325–332.
10. Hilfiker, H., Hattenhauer, O., Traebert, M., Forster, I., Murer, H., and Biber, J. (1998). Characterization of a murine type II sodium-phosphate cotransporter in mammalian small intestine. *Proc. Natl. Acad. Sci. USA.* **95:** 14564–14569.
11. Kos, C. H., Tihy, F., Murer, H., Lemieux, N., and Tenenhouse, H. S. (1996). Comparative mapping of Na^+-phosphate cotransporter genes, NPT1 and NPT2, in human and rabbit. *Cytogenet. Cell. Genet.* **75:** 22–24.
12. Magagnin, S., Werner, A., Markovich, D., Sorribas, V., Stange, G., Biber, J., and Murer, H. (1993). Expression cloning of human and rat renal cortex Na/P_i cotransport. *Proc. Natl. Acad. Sci. USA.* **90:** 5979–5983.
13. McPherson, J. D., Krane, M. C., Wagner-McPherson, MzfnC. B., Kos, C. H., and Tenenhouse, H. S. (1997). High resolution mapping of the renal sodium-phosphate cotransporter gene (NPT2) confirms its localization to human chromosome 5q35. *Pediatr. Res.* **41:** 632–634.
14. Miyamoto, K., Tatsumi, S., Sonoda, T., Yamaoto, H., Minami, H., Taketani, Y., and Takeda, E. (1995). Cloning and functional expression of a Na^+-dependent phosphate co-transporter from human kidney: cDNA cloning and functional expression. *Biochem. J.* **3005:** 81–85.
15. Nakagawa, N., and Ghishan, F. K. (1994). Low phosphate diet upregulates the renal and intestinal sodium-dependent phosphate transporter in vitamin D-resistant hypophosphatemic mice. *Proc. Soc. Exp. Biol. Med.* **205:** 162–167.
16. Shibui, A., Tsunoda, T., Seki, N., Suuzuki, Y., Sugane, K., and Sugano, S. (1999). Isolation and chromosomal mapping of a novel human gene showing homology to Na^+/PO_4 cotransporter. *J. Hum. Genet.* **44:** 190–192.

IX D. 20058c Applicants' copy 6x Ref 4

Predicting functions from protein sequences—where are the bottlenecks?

Peer Bork[1] & Eugene V. Koonin[2]

The exponential growth of sequence data does not necessarily lead to an increase in knowledge about the functions of genes and their products. Prediction of function using comparative sequence analysis is extremely powerful but, if not performed appropriately, may also lead to the creation and propagation of assignment errors. While current homology detection methods can cope with the data flow, the identification, verification and annotation of functional features need to be drastically improved.

With the rapid growth of sequence-related and other databases, there is increasing concern about the impact of this information explosion[1,2]. Is the burgeoning diversity of information an advantage or will it play havoc with genome analysis and ultimately lead to an error catastrophe? The eventual success of genome projects will depend on our ability to handle information in a manner that enhances the capability for function prediction rather than pollutes the analysis with noise. Only a minority of sequenced genes have been studied in direct experiments. In the foreseeable future, the gap between the number of sequences available and the extent of functional characterization of gene products is expected to broaden even further. It is apparent that computer procedures for the prediction of functional features from sequence are much faster and cheaper than 'wet' experiments and, by default, are applied to each gene that is sequenced. This puts tremendous pressure on computational approaches to ascribe as much functional information as possible to each gene. It appears, however, that within the typical framework of current sequencing projects, optimization of the computer analysis and functional annotation has not yet been achieved. A testimony to this is the repeated discovery of new, functionally relevant features in sequences that already have been subjected to standard computer procedures.

Given the database growth and the accompanying increase in noise and redundancy, we believe that there are currently two major bottlenecks that need to be overcome *en route* to efficient functional predictions from protein sequences. First, there is the lack of a widely accepted, robust and continuously updated suite of sequence analysis methods integrated into a coherent and efficient prediction system. Second, there is considerable 'noise' in the presentation of experimental information, leading to insufficient or erroneous functional assignment in sequence databases.

Here we review some computer-based approaches that allow utilization of more functional and structural information than the current standard schemes, and discuss some of the difficulties in handling and interpreting functional information.

Effects of database growth

From a purely statistical standpoint, the chances of detecting significant similarity between a new sequence and the ones already available in databases decrease with the expansion of the search space, in this case, database growth[3,4]. Fortunately, at least three major factors counter this adverse statistical effect. First, the sequence space is not infinite: new sequences fill it and inevitably increase the chance of finding homologues. Second, complete genome sequences of phylogenetically distant species bring a qualitative improvement to the representation of conserved gene families[5]. With numerous genome sequences of unicellular organisms already available and the majority of human genes represented in the Expressed Sequence Tags (EST) database[6], it is becoming increasingly likely that a family to which a new protein belongs is already represented in the databases[7]. Third, the development of new, more sensitive methods for information filtering and database searching, as well as improved strategies for their application, result in the delineation of previously undetectable, subtle relationships between sequences.

The net effect is that, for a given sequence, the likelihood of detecting a homologue in the databases steadily increases with time. To illustrate this, we followed the kinetics of homology identification and function prediction for an unbiased data set, namely the proteins encoded by the genes on yeast chromosome III. After the initial characterization of this eukaryotic chromosome in 1992 (ref. 8), and early efforts to push the limits of computer-aided predictions[9–11], there has been a continuous linear increase in the fraction of proteins that have identifiable homologues and predictable functions (Fig. 1). Although due in part to a decrease in the number of open reading frames (ORFs) identified as likely genes, this trend demonstrates the increasing utility of computer analysis despite database growth. Of the current set of 25 predicted genes without homology or functional assignment, 15 are smaller than 150 amino acids and may not be expressed at all[12]. With homologues now detectable for 85% of the proteins encoded by genes on yeast chromosome III and at least some functional features identified for 70%, we may soon approach an upper limit for computer-aided predictions. The depth of the functional characterization attainable for many proteins, however, will continue to increase, which is not reflected in the above numbers.

Reducing the noise in sequence searches

In-depth analysis of protein sequences often results in functional predictions not attained in the original studies. This can be illustrated by the results obtained with genes mutated in human diseases (disease genes). Identification of such genes typically involves the time- and labour-consuming process of positional cloning[13]. It is therefore critical that as much functionally relevant information as possible is extracted from the protein sequence encoded by a disease gene once it becomes available. It is not uncommon, however, that rapid computer re-analysis produces unexpected insight into the evolutionary relationships and

[1]*EMBL, Meyerhofstr. 1, 69012 Heidelberg and Max-Delbrück-Center for Molecular Medicine, Berlin-Buch, Germany.* [2]*NCBI, NIH, Bethesda, USA. Correspondence should be addressed to P.B. email: bork@embl-heidelberg.de*

Table 1• Selected examples of computer-aided discoveries with positionally cloned disease genes

Gene (protein)	Phenotype/ Disease	Original observation on protein sequence	New findings by in silico computer analysis and implications	Subsequent experimental support
LEP (leptin)	Hereditary obesity	None[23]	Structural similarity to helical cytokines identified by threading; likely cytokine activity and features[24]	Leptin structure is typical of helical cytokines[25]; leptin receptor is homologous and functionally analogous to cytokine receptors[26]
DMD (dystrophin)	Muscular dystrophy	Spectrin repeats[40]	WW, and ZZ signalling domains[41,42]	3D structure of WW and ligand[43]
HD (huntingtin)	Huntington disease	None[44]	HEAT repeats covering a considerable fraction of the protein and indicating structural features[45]	Conservation of repeats in homologues
BRCA1 (BRCA1)	Hereditary breast cancer	N-terminal RING-finger domain[46]	C-terminal BRCT domain conserved in many DNA repair-dependent checkpoint proteins; likely role in cell cycle checkpoints[47–49]	BRCT domain coincides with transcription activation domain[50], BRCA1 involved in repair and cell cycle [51]
BRCA2 (BRCA2)	Hereditary breast cancer	Coiled-coil domains[52]	Previously unknown repeats covering almost a third of this large protein[53]	Verification of the repeats by sequencing homologues[54]
CHM (CHM)	Choroideremia (hereditary blindness)	Homologue of guanine nucleotide dissociation inhibitor Rab-GDI[55]	FAD (NAD)-binding domain[56]	3D structure confirmed presence of a dinucleotide-binding domain[57]
FRDA1 (frataxin)	Freidrich ataxia	Highly conserved eukaryotic homologues[58]	Bacterial homologues (CyaY); on the basis of phylogenetic analysis, a mitochondrial function suggested for frataxin[59]	Mitochondrial localization of frataxin demonstrated[60]
CLCN1 (CLCN1)	Myotonia (Thomsen disease)	Chloride channel[61]	CBS domain[62]	NA
TAZ (tafazzin)	Barth syndrome	None[63]	Acyltransferase domain; possible role in membrane biogenesis[64]	NA
MLH1 (MLH1)	Hereditary non-polyposis colon cancer	Homologue of bacterial and yeast DNA repair protein MutL[65]	ATPase domain conserved in topoisomerase type II (3D structure available), HS90, and His kinases; predicted ATPase activity[14,66]	NA
WRN (WRN)	Werner syndrome (premature aging)	DNA helicase domain (RecQ homologue)[67]	N-terminal exonuclease domain (structure known for homologous domain in bacterial PolA) and putative C-terminal RNA-binding domain conserved in BLM and RNA-ase D; predicted exonuclease activity[14,68,69]	NA
BLM (BLM)	Bloom syndrome	DNA helicase domain (RecQ homologue)[70]	C-terminal nucleic acid-binding domain conserved in WRN and RNAase D[69]	NA
WAS (WAS)	Wiskott-Aldrich syndrome	WH1 (ref. 71)	WH1 in homer with ligand-binding implications[72]	NA
SCA2 (ataxin-2)	Spinocerebral ataxia-2	Polyglutamine stretch expanded in ataxia[73]	Novel conserved domain shared with spliceosomal Sm proteins and bacterial global transcription regulators; possible role in splicing[74]	NA

NA, no data available

likely functions of disease gene products (see Table 1 for selected examples, ref. 14 for a more detailed list). These predictions have the potential to greatly facilitate and accelerate functional characterization of disease genes, and ultimately, the development of therapeutics. Thus, it is instructive to explore the main reasons why these relationships have been missed in the original studies but have been revealed subsequently.

Generally, the problem of obtaining the best results from a database search can be formulated in terms of signal-to-noise ratio[15]. One way to compensate for noise is to increase the sensitivity of the search method. The recently developed PSI-BLAST method (Position-Specific Iterative BLAST; ref. 16) is the latest major step in this direction. It combines the advanced version of the popular BLAST algorithm, which has been modified to incorporate gapped alignments, with profile analysis. PSI-BLAST is fast and highly sensitive and, when combined with appropriate filtering of low complexity regions (see below), also highly selective. Nevertheless, there is still room for improvement as the signal-to-noise ratio can be increased further using protein-family–dependent strategies such as concentrating on motifs, local conserved regions (that is, lowering the signal, but to a lesser extent than the noise) or including global alignments to increasing the signal (but to a greater extent than the noise, due to variable regions; ref. 15).

To reduce the noise, it is advisable first to pre-process the query sequence to account for the different types of compositionally-biased regions. This is achieved by filtering low-complexity regions using the SEG program[17], which is currently implemented as a default parameter in the BLAST programs. Approximately 15% of protein-sequence databases appear to be represented by these low complexity regions, which are thought to correspond to non-globular domains; most of these are found in eukaryotic proteins[18]. Low-complexity sequences tend to produce database hits with artifactually low *P*-values (estimated probability of a random match) simply because there are many sequences with a similarly reduced residue alphabet in the database. For the detection of particular biases such as coiled-coil regions, accurate methods have been developed[19] that also should be included in a pre-processing step. In addition, programs exist for delineating other low-complexity regions such as transmembrane segments[20], although one should always be aware of the prediction accuracy and the choice of parameters that need to be adapted to the specific problem (Table 2). These *caveats* notwithstanding, integration of different pre-processing methods into the search scheme considerably increases the signal-to-noise ratio.

Noise reduction is particularly important when dealing with multidomain proteins. Usually each of the domains not only is a defined structural unit with its own evolutionary history but also carries a distinct subfunction contributing to a more complex overall function of the protein. In a standard database search, the multidomain architecture of proteins causes several problems that frequently hamper extraction of functional information from sequences. First, the domain with the strongest signal will always score highest so that lower-scoring domains in different locations in the query sequence may be overlooked. Second, hits that involve single domains are often misleading with regard to functional characterization. For example, identification of a src-homology domain (SH3) of a src-like protein kinase within a query sequence has often caused mis-annotation of the query as a protein kinase. Third, searching with the entire sequence of a multidomain protein is much less sensitive than searching with segments that are located between known domains. Thus, scanning databases of known domains such as PROSITE, BLOCKS, PRINTS, PFAM or SMART (refs 21,22) is an important complement to standard database searches.

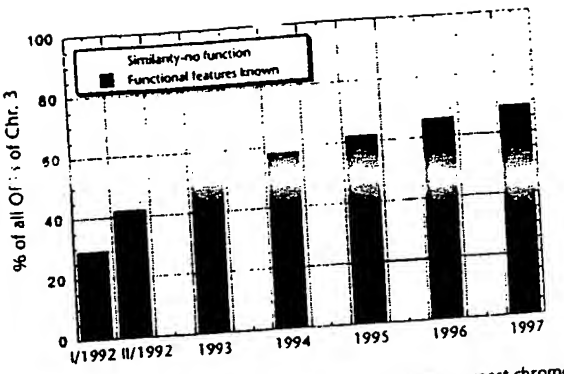


Fig. 1 Increase in knowledge of ORFs encoded by genes on yeast chromosome III, as recorded at the end of each year. Note that the number of ORFs (originally 182; ref. 8) is steadily decreasing and might fall below 150 because 16 of the 25 remaining ORFs without homologues are small and have unusual amino-acid compositions. On the other hand, an ORF with a homologue in another species (85% of the total) is very likely to encode an expressed protein. The term 'function' is loosely defined and we have included even weak functional features such as the presence of an SH3 domain or a defined ATP-binding motif, which hint at cellular or molecular roles. There are currently only three proteins encoded by genes on chromosome III which have a described function but lack protein homology.

Multidomain architecture is most common in eukaryotic proteins—and the majority of disease gene products contain multiple domains[14,22]. Indeed, computational re-analysis has revealed numerous undetected domains in disease genes, which suggest additional functional and structural features (Table 1).

In the absence of recognizable sequence similarity, 'threading' approaches—that is, fold assignments by means of checking for sequence compatibility with known three-dimensional structures—might reveal additional structural insights, as has been successfully demonstrated for leptin, the obesity gene product[23]. The prediction of a helical cytokine-like fold[24] in leptin has been confirmed both by direct structure determination[25] and by numerous functional studies[26]. However, the accuracy of threading methods is limited[27,28], and the sensitivity of the most powerful modern methods for sequence similarity analysis approaches that of threading. For example, highly accurate fold assignments for more than 35% of all *Mycobacterium genitalium* proteins (or some of their domains) are currently possible using iterative homology searches (Huynen, M.A. *et al.*, submitted). This fraction includes the 8% which previously has been claimed to be identifiable only by using threading techniques[29].

Effects of noise on functional predictions

Functional information is hard to quantify and a critical point in an analysis is determining how many functional features are shared between the protein in question and sequence(s) similar to it in the database. In the best case, they are orthologues[30,31], corresponding genes in two species that perform the same molecular function. Even this, however, is often difficult to discern in a situation where the database hits are paralogues (other members of a multigene family which have distinct, but perhaps related, functions). With many complete genome sequences available, databases of orthologues are expected to become indispensable for functional annotation[5]. However, more often than not, it is clear that the cellular role of the protein in question differs from that of the detected homologue(s) and there is currently no automatic means to establish how much functional information can be legitimately transferred by analogy from the homologue to the query[1].

As the problem of orthology *versus* paralogy has only recently been introduced into genome analysis, many erroneous func-

Table 2 • A checklist for in-depth analysis of protein sequences and prediction of function from sequence[a]

Procedure	Purpose and comment
	Identification of and filtering for structural features
Mask non-globular or highly compositionally-biased regions (reduced residue alphabet)	Reduce noise and avoid spurious hits due to low complexity regions
Mask coiled-coil regions	This is a special type of low-complexity region that causes numerous other coiled-coil regions to match but is not efficiently detected by general methods for complexity analysis
Identify transmembrane regions (including signal sequences and GPI anchors)	Yet another form of composition bias that may result in matches with non-homologous membrane regions; the presence of these domains should be taken into account in all functional predictions.
Identify internal repeats	Reduces the search space for remaining parts and also may lead to the detection of novel repeat types
Predict secondary structure	The best programs predict a protein's structural class; use of multiple alignments significantly improves the accuracy of prediction
	Identification of homologues
Identify known domains in dedicated databases (for example, Pfam, PROSITE, BLOCKS, PRINTS, SMART) prior to a BLAST search	Identification of annotated domains may be more sensitive when these databases are used; removal of known domains also reduces the search space for remaining parts of the protein.
Search complete sequence databases with subsequences of long (>200 a.a.) proteins individually; preferably use subsequences separated by known domains or low-complexity regions	Increase search sensitivity by reducing the search space; exclude domains with numerous homologues (for example, protein kinases), which may obscure even highly significant similarity to other domains of the query
Perform reciprocal searches to verify weak similarity to possible homologues	The alignments of a potentially relevant database hit with its indisputable homologues support (if the conservation pattern is consistent) or reject (if it is different) a weak pairwise similarity.
Perform exhaustive, iterative database searches	Database search methods are non-transitive and non-symmetric; therefore analysis of a protein family should be performed iteratively, starting with different members, until no new homologues are detected
Combine search for pairwise sequence similarity (for example, first BLAST scan) with profile, motif, and pattern searches (explicit example in Psiblast[16]) or by using the various programs available[15]	The information contained in a multiple alignment provides for amplification of weak but potentially important sequence signals and is indispensable for the delineation of protein superfamilies.
	Prediction of protein functions
Carefully consider domain organization and distinct functions of individual domains	Many proteins are multifunctional; assignment of a single function, which is still common in genome projects, results in loss of information and outright errors
Do not take database annotation for granted, especially if only one homologue is detectable or there is inconsistency between different homologues	Databases contain a number of incorrect annotations due to experimental errors as well as functional assignments on the basis of dubious sequence similarity
Do not simply transfer functional information from the best hit	The best hit is frequently hypothetical or poorly annotated; other hits with similar or even lower scores may be more informative; even the best hit may have a different function (see below)
Do cluster analysis of the homologues to identify the appropriate level of precision for functional prediction	It is typical that the general function of a protein can be identified easily but the prediction of substrate specificity is unwarranted; for example, many permeases of different specificity show approximately the same level of similarity to each other
Check sequence context (e.g. likely clashes in the co-occurrence of a signal sequence and a zinc-finger or glycolysation sites in a cytoplasmic protein)	Comparison of different predicted structural and functional features helps avoiding erroneous predictions
Identify similarities to proteins with known 3D structure	Models of highly conserved homologues can be built and might reveal further functional insights

[a]Complementary checklists can be found in ref. 15

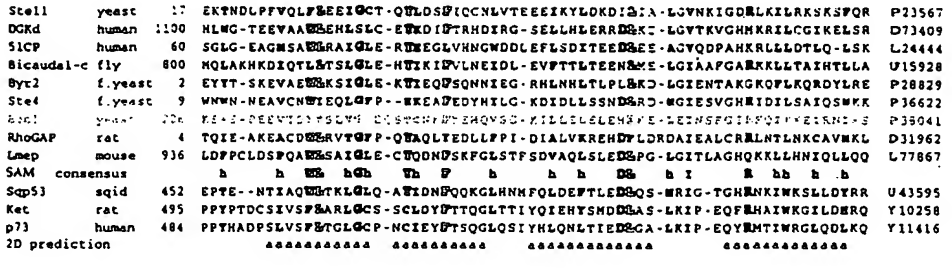


Fig.2 Sequence alignment of a selected set of SAM domains with p73-like proteins. Conserved hydrophobic positions are shown in bold; residues that are conserved in at least 75% of the sequences are highlighted. The SAM consensus[33] corresponds to the consensus of the alignment. Secondary structure predictions ('a' represents 'alpha helix') have been taken from ref. 33. Position in the sequence and database accession numbers are given in the third and last column, respectively.

tional annotations have already been incorporated and subsequently propagated in sequence databases. A quality index for functional annotation in databases still remains a distant goal and new approaches are required to improve the sensitivity of functional characterization so as to avoid functional over- and under-predictions for a given database match.

Transfer of functional information from sequences in the database to the query is also hampered by the effects of noise in the functional description of proteins. Updates on functional features require an awareness of the scientific literature; an experiment in one species on a previously sequenced gene brings important consequences for homologous genes in other species. This has led to a gap between functional information contained in the sequence databases and the specialized knowledge embodied in the literature. At present, there is no automatic method that can replace literature searches. A recent case in point is p73, a human paralogue of the tumour suppressor p53. It contains a carboxy-terminal extension, for which similarity has only been found in squid 'p53' (ref. 32). However, relevant and potentially important information about this region could be obtained by simply searching the PubMed database for the combination of terms 'p53' and '*Loligo*' (Latin for squid). It has been shown that the squid p53 homologue contains a C-terminal SAM domain, a distinct protein-protein interaction and dimerization domain found primarily in developmental regulators[33]. Furthermore, just weeks after the publication of human p73, the gene encoding rat KET protein, a close p73 homologue was sequenced[34]. With the KET sequence in the database, a PSI-BLAST search using the conservation profile of the two p73 species readily reveals significant similarity between their C-terminal regions and numerous SAM domains (Fig. 2). The SAM domain in p73 may be involved in dimerization and may mediate an interaction with another protein(s) involved in transcription regulation and developmental control of gene expression. A more thorough literature search or the use of 'awareness' tools (see below) would have helped to retrieve, in an automatic manner, potentially important information on p73.

Toward integration of functional and structural features

In summary, the currently available methods for sequence analysis are sophisticated, and while further improvements will certainly ensue, they are already capable of extracting subtle but functionally relevant signals from protein sequences. Whether or not a researcher actually reaps maximal benefit from such analysis, however, depends on the application of an appropriate combination of methods in the correct setting. There is no single, universal recipe for this purpose, but we have attempted to compile a short checklist, which will minimize the risk of missing important functional signals hidden in protein sequences (Table 2).

A comprehensive, precisely defined and standardized classification of biological functions is required for the automation of the prediction of gene functions. The task of constructing such a classification is immensely difficult given that even small proteins—not to mention large, multidomain ones—are certain to have multiple roles in the cell. Classification schemes that have been proposed for prokaryotic gene products[35–37] are useful in comparative genome analysis but grossly oversimplify the problem. We believe that, at present, there is still no proper language for the adequate and uniform description of functions and therefore it is hard to predict when, if at all, a (a)periodic system of biological functions may become a reality.

Generally, the incorporation of known functional information into databases at various levels is a pressing need, requiring the combined efforts of experimentalists, computational biologists and database developers. The challenge is particularly formidable as new types of information, such as tissue- and organ-specific gene expression patterns on a genome scale as well as numerous data on protein interactions and post-translational modifications are rapidly becoming available (for details see refs 38,39).

This continuous flow of information also requires 'update' and 'awareness' tools that filter and incorporate incoming data (sequences, literature, *etcetera*) and new applications (servers, methods). Ideally, these tools will notify researchers upon subscription using a customized query profile (keywords, sequences of interest, problem description) or systematically integrate the filtered information into dynamic databases. The major problem that remains is the quality of the information, as no data-mining tools yet exist that can judge the various experimental protocols described in the respective literature.

There is little doubt that a new generation of bioinformatics approaches will soon integrate sequence and structure analysis methods with awareness tools, information filters and dynamic data processing in preparation for the forthcoming postgenomics aera. However, all these tools will only facilitate but not replace the work of a scientist who defines the questions and interprets the results.

Note added in proof: Elizabeth Greene & Steven Henikoff have constructed an article for the *Nature Genetics* website (http://www.genetics.nature.com/gazing/), which provides direct links to a range of web-based tools for database sequence searches, and homology and structural predictions for query protein sequences.

1. Bork, P. & Bairoch, A. Go hunting in sequence databases but watch out for the traps. *Trends Genet.* **12**, 425–427 (1996).
2. Bhatia, U., Robison, K. & Gilbert, W. Dealing with database explosion: a cautionary note. *Science* **276**, 1724–1725 (1997).
3. Altschul, S.F., Boguski, M.S., Gish, W. & Wootton, J.C. Issues in searching molecular sequence databases. *Nature Genet.* **6**, 119–129 (1994).
4. Smith, R.F. Sequence database searching in the era of large scale genomic sequencing. *Genome Res.* **6**, 653–660 (1996).
5. Tatusov, R.L., Koonin, E.V. & Lipman, D.J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
6. Boguski, M.S., Tolstoshev, C.M. & Bassett, D.E. Jr. Gene discovery in dbEST. *Science* **265**, 1993–1994 (1994).
7. Green, P., Lipman, D., Hillier, L., Waterstone, R., States, D. & Claverie, J.-M. Ancient conserved regions in new gene sequences and the protein databases. *Science* **259**, 1711–1716 (1993).
8. Oliver, S.G. *et al.* The complete sequence of yeast chromosome III. *Nature* **357**, 38–46 (1992).
9. Bork, P. *et al.* What's in a genome? *Nature* **358**, 287 (1992).
10. Sharp, P.M. & Lloyd, A.T. Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. *Nucleic Acids Res.* **21**, 179–183 (1993).
11. Koonin, E.V., Bork, P. & Sander, C. Yeast chromosome III: New gene functions. *EMBO J.* **13**, 493–503 (1994).
12. Fickett, J.W. ORF's and genes: how strong a connection? *J. Comput. Biol.* **2**, 117–123 (1995).
13. Collins, F. S. Positional cloning from moves from perditional to traditional. *Nature Genet.* **9**, 347–350 (1995).
14. Mushegian, A.R., Bassett, D.E. Jr, Boguski, M., Bork, P. & Koonin, E.V. Positionally cloned human disease genes: Patterns of evolutionary conservation. *Proc. Natl. Acad. Sci. USA* **94**, 5831–5836 (1997).
15. Bork, P. & Gibson, T.J. Applying motif and profile searches. *Methods Enzymol.* **266**, 162–184 (1996).
16. Altschul, S.F *et al.* Gapped Blast and PSI-Blast, a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
17. Wootton, J.C. & Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266**, 554–571 (1996).
18. Wootton, J.C. Sequences with unusual amino acid composition. *Curr. Opin. Struct. Biol.* **4**, 413–421 (1994).
19. Lupas, A. Predicting coiled coil regions in proteins *Curr. Opin. Struct. Biol.* **7**, 388–393 (1997).
20. Rost, B. & O'Donoghue, S. Sysisphus and the prediction of protein structure *Comput. Appl. Biosci.* **13**, 345–356 (1997).
21. Henikoff, S. *et al.* Gene families: the taxonomy of protein paralogues and chimeras. *Science* **278**, 609–613 (1997).
22. Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P. SMART, a simple modular architecture research tool: Identification of signalling domains *Proc. Natl. Acad. Sci. USA*, in press.
23. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
24. Madej, T., Boguski, M.S. & Bryant, S.H. Threading analysis suggests that the obese gene product may be a helical cytokine. *FEBS Lett.* **373**, 13–18 (1995).
25. Zhang F. *et al.* Crystal structure of the obese protein leptin-E100. *Nature* **387**, 206–209 (1997).
26. Tartaglia, L.A. The leptin receptor. *J. Biol. Chem.* **272**, 6093–6096 (1996).
27. Rost, B., Schneider, R. & Sander, C. Protein fold recognition by prediction-based threading. *J. Mol. Biol.* **270**, 471–480 (1997).
28. Smith, T.F. *et al.* Current limitations to protein threading approaches. *J. Comput. Biol.* **4**, 217–225 (1997).
29. Fischer, D. & Eisenberg, D. Assigning folds to the proteins encoded by the genome of *Mycoplasma genitalium*. *Proc. Natl. Acad. Sci. USA* **94**, 11929–11934 (1997).
30. Fitch, W.M. Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99–113 (1970).
31. Fitch, W.M. Uses for evolutionary trees. *Phil. Trans. R. Soc. Lond. B.* **349**, 93–102 (1995).
32. Kaghad, M. *et al.* Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90**, 809–819 (1997).
33. Schultz, J., Ponting, C.P., Hofmann, K. & Bork, P. SAM as a protein interaction domain involved in developmental regulation. *Prot. Sci.* **6**, 249–253 (1997).
34. Schmale, H. & Bamberger, C. A novel protein with strong homology to the tumor surpressor p53. *Oncogene* **15**, 1363–1367 (1997).
35. Riley, M. Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* **57**, 862–952 (1993).
36. Bork, P. *et al.* Exploring the *Mycoplasma capricolum* genome: a minimal cell reveals its physiology. *Mol. Microbiol.* **16**, 955–967 (1995).
37. Tatusov, R.L. *et al.* Metabolism and evolution of *Haemophilus influenzae* deduced from a whole genome comparison to *Escherichia coli*. *Curr. Biol.* **6**, 279–291 (1996).
38. Hieter, P. & Boguski, M. functional genomics: It's all how you read it. *Science* **278**, 601–602 (1997).
39. Zhang, L. *et al.* Gene expression profiles in normal and cancer cells. *Science* **276**, 1268–1272 (1997).
40. Koenig, M., Monaco, A.P. & Kunkel, L.M. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* **53**, 219–226 (1988).
41. Bork, P. & Sudol, M. The WW domain: a signalling site in dystrophin? *Trends Biochem. Sci.* **19**, 531–533 (1994).
42. Ponting, C.P., Blake, D.J., Davies, K.E., Kendrick-Jones, J. & Winder, S.J. ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins. *Trends Biochem. Sci.* **21**, 11–13 (1996).
43. Macias, M.J. *et al.* Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**, 646–649 (1996).
44. Huntington's disease collaborative research group. A novel gene containing a trinucleotide repeat that is expanded and unstable in huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
45. Andrade, M. & Bork, P. HEAT repeats in Huntington's disease protein. *Nature Genet.* **11**, 115–116 (1995).
46. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
47. Koonin, E.V., Altschul, S. & Bork, P. BRCA1 protein products: functional motifs. *Nature Genet.* **13**, 266–268 (1996).
48. Bork, P. *et al.* A superfamily of conserved domains in DNA damage responsive cell cycle checkpoint proteins. *FASEB J.* **11**, 68–76 (1997).
49. Callebaut, I. & Mornon, J.P. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400**, 25–30 (1997).
50. Monteiro, A.N.A., August, A. & Hanafusa, H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. *Proc. Natl. Acad. Sci. USA* **93**, 13595–13599 (1996).
51. Scully, R. *et al.* Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage. *Cell* **90**, 425–435 (1997).
52. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792 (1995).
53. Bork, P., Blomberg, N. & Nilges, M. Internal repeats in the BRCA2 protein sequence. *Nature Genet.* **13**, 22–23 (1996).
54. Bignell, G., Micklem, G., Stratton, M.R., Ashworth, A. & Wooster, R. The BRC repeats are conserved in mammalian BRCA2 proteins. *Hum. Mol. Genet.* **6**, 53–58 (1997).
55. Cremers, F.P. *et al.* Cloning of a gene that is rearranged in patients with choroideraemia. *Nature* **347**, 674–677 (1990).
56. Koonin, E.V. Human choroideraemia protein contains an FAD-binding domain. *Nature Genet.* **12**, 237–239 (1996).
57. Wu, S.K., Zeng, K., Wilson, I.A. & Balch, W.E. Structural insights into the function of the Rab GDI superfamily. *Trends Biochem. Sci.* **21**, 472–476 (1996).
58. Campuzano, V. *et al.* Freidrich's ataxia: autosomal recessive disease caused by an intronic GAA repeat expansion. *Science* **271**, 1423–1427 (1996).
59. Gibson, T., Koonin, E.V., Musco, G. Pastore, A. & Bork, P. Freidrich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction. *Trends Neurosci.* **19**, 465–468 (1996).
60. Koenig, M. & Mandel, J.-L. Deciphering the cause of Freidrich's ataxia. *Curr. Opin. Neurobiol.* **7**, 689–694 (1997).
61. Koch, M.C. *et al.* The skeletal muscle chloride channel in dominant and recessive myotonia. *Science* **257**, 797–800 (1992).
62. Bateman, A. The structure of a domain common to archebacteria and the homocystinuria disease protein. *Trends Biochem. Sci.* **22**, 12–13 (1997).
63. Bione, S. *et al.* A novel X-linked gene, G4.5 is responsible for Barth syndrome. *Nature Genet.* **12**, 385–389 (1996).
64. Neuwald, A.F. Barth syndrome might be due to acyltransferase deficiency. *Curr. Biol.* **7**, 465–466 (1997).
65. Kolodner, R. *et al.* Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev.* **10**, 1433–1442 (1996).
66. Bergerat, A. *et al.* An atypical topoisomerase II from Archea with implications for meiotic recombination. *Nature* **386**, 414–417 (1997).
67. Yu, C.E. *et al.* Positional cloning of the Werner's syndrome gene. *Science* **272**, 258–262 (1996).
68. Mian, I.S. Comparative sequence analysis of ribonucleases HII, II, II PH and D. *Nucleic Acids Res.* **25**, 3187–3195 (1997).
69. Morozov, V., Mushegian, A.R., Koonin, E.V. & Bork, P. A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases. *Trends Biochem. Sci.* **22**, 417–418 (1997).
70. Ellis, N.A. *et al.* The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**, 655–666 (1995).
71. Symons, M. *et al.* Wiskott-Aldrich syndrome protein, a novel effector for the GTPase CDC42Hs, is implicated in actin polymerization. *Cell* **84**, 723–734 (1996).
72. Ponting, C.P. & Phillips, C. Identification of homer as a homologue of the Wiskott-Aldrich syndrome protein suggests a receptor-binding function for WH1 domains. *J. Mol. Med.* **75**, 769–771 (1997).
73. Imbert, G. *et al.* Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nature Genet.* **14**, 285–291 (1996).
74. Neuwald, A.F. & Koonin, E.V. Ataxin-2, global regulators of bacterial gene expression, and spliceosomal snRNP proteins share a conserved domain. *J. Mol. Med.* **76**, 3–5 (1998).

IX E. R0058C Applicants' copy Ex. Ref. V

BIOINFORMATICS

Vol. 14 no. 9 1998
Pages 753-754

Editorial

What We Do Not Know About Sequence Analysis and Sequence Databases

The marriage of high-throughput nucleotide sequencing with computational methods for the analysis of nucleotide and protein sequences have ushered in a new era of molecular biology. Entire genomes are deposited into the sequence DBs at a growing rate. Typically, investigators can use computational sequence analysis to assign functions to the majority of the open reading frames in genome sequences. That analysis can identify a surprisingly large fraction of the genes within the organism. That fraction is increasing over time as the sequence databases contain a larger fraction of all functional domains.

The growing wealth of information within the sequence databases provides a foundation for the biology of the 21st Century. We will mine these data for decades to come, developing complex and incredibly accurate cellular models that can predict the behavior of living systems by integrating across the functions of their molecular parts.

Or will we? Although the preceding scenario is the likely one, we would be irresponsible to not consider another possible outcome: an explosion of incorrect annotations within the sequence databases. Each new sequence deposited in the public databases has been annotated with respect to those same databases. Functional annotations are propagated repeatedly from one sequence to the next, to the next, with no record made of the source of a given annotation, leading to a potential transitive catastrophe of erroneous annotations. Investigators who later attempt to separate the wheat from the chaff will discover that they cannot simply retreat to the safety of experimentally annotated sequences by ignoring the computationally annotated sequences, because the public DBs do not explicitly distinguish the two sets. In fact, the public sequence DBs keep virtually no tracking information about the methods used to annotate their data.

Can we rule this possibility out on any objective grounds? No. We have no reliable data regarding either the current rate of errors (incorrect functional annotations) within the public DBs, nor on the rate of change of that error rate (we do not even know if it is increasing or decreasing each year).

Many years of research have led to the development of detailed statistical models for sequence-similarity searching algorithms such as FASTA and the BLAST family of programs. Researchers employ these algorithms to identify the functions of novel sequences in two phases. In phase I, they identify homologs of a novel sequence. In phase II, they infer the function of the novel sequence with respect to the homologs that have been identified by examining the range of functions of the homologs and the sequence regions those homologs share with the novel sequence. A number of studies have examined the sensitivity and specificity with which those algorithms identify homologs of a query sequence in a sequence database (phase I). However, few if any studies have evaluated the accuracy of phase II, or of the complete functional-prediction process. Incorrect functional predictions can result from a number of causes, including: divergence of function within homologous proteins; confusion or omission of functions across multimodular proteins (e.g. if only one function of a multimodular protein is given in its description line, but a region of the protein containing a different function is what matched the query sequence); and omission of phase II altogether by simply choosing the strongest homolog as the source of attributed function.

Although we know the accuracy with which sequence homologs can be determined, we know little about the accuracy of the overall process of assigning function by homology. Thus, remarkably, just as we have no idea what is the rate of errors in the database foundation of biology for the 21st century, we also have little hard information about the accuracy of the most commonly used computational method in bioinformatics — the method that underpins all genome sequence analysis.

Consider the following additional questions. It is likely that different people carry out phase II with somewhat different methods that produce different accuracies: what is the range of accuracies? In the past few years, however, programs such as GeneQuiz and Magpie have begun to tackle phase II. What are their accuracies? Are the programs more or less accurate than expert scientists? To address these questions it would help greatly if a sequence-analysis benchmarking test-suite were available, i.e., a set of sequences whose functions had been established experimentally and were known to be correct, and that could be used as a test set for evaluating programs or scientists. (Such test suites have been of great value in the protein-structure prediction field.) To critically evaluate these programs and to learn from their successes and failures, it will prove crucial for their authors to publish the decision rules that the programs employ. However, these rules have typically not been published in past articles. Although a number of new methods have recently been proposed for automated sequence analysis (such as the COGS method), it is hard to see why we should have any confidence in the claims made about the accuracy of these methods when those claims are based on a handful of examples rather than on systematic empirical studies.

To assemble a set of sequences for a sequence-analysis benchmark, we would attempt to find a set of sequences whose functions are thought to be correct with high confidence. Unfortunately, the current sequence DBs provide little assistance in finding such a set of sequences. We have more faith in the correctness of those sequences whose functions were determined experimentally, rather than through computational means. But the sequence DBs do not

distinguish experimentally determined functions from those determined computationally, much less do they associate a level of confidence with each functional assignment. (Note that Swiss-Prot does distinguish some sequence features as 'from sequence analysis', but not the overall function assigned to the sequence.)

If we had reason to question the accuracy of a particular functional assignment in a sequence DB, what would we want to know about the sequence? We would want to know if the function was determined experimentally or computationally; if computationally, we would want to know whether a person or a program made the final decision. If a program, we would want to know which program; if a person, we would want to know which similarity-searching program (or set of programs) the person relied on. We would also like to know when the function was assigned (was the assignment based on up-to-date computational techniques and up-to-date databases?). We would also like to know the overall rate of errors in the functional assignments in the sequence DB we are looking at.

It is extremely difficult both to estimate the error rates of any sequence DB as a whole, and to estimate the reliability of any particular entry within these DBs, for three principal reasons.

First, the DBs themselves provide very little metadata — historical or tracking data ABOUT the primary data within the DB (such as a level of confidence in a functional annotation, or the name of the program that created the functional annotation). Without such metadata it will be difficult for scientists — and virtually impossible for programs — to make intelligent decisions about what data to trust.

Second, the sequence DBs typically do not publish detailed descriptions of their methodologies to tell us exactly what manual and automated procedures they subject each sequence to. For example, we do not know what checking or correction procedures new GenBank entries are subjected to. We also do not know precisely what sequence-analysis procedures are used to annotate Swiss-Prot or TREMBL. These DBs do not simply accept sequences annotated by other scientists — they perform sizeable annotation operations of their own which should be documented in detail. We also do not know the rate at which wrong annotations in GenBank or Swiss-Prot are corrected. The rates are probably very different because Swiss-Prot accepts corrections from its user community, whereas GenBank apparently only accepts corrections from the author of an entry. We also do not know the error rate in the corrections! The procedures used by each DB essentially tell us 'default metadata' for each DB entry. In fairness, one reason such descriptions are not published is that journals have apparently not accepted descriptions of these procedures that have been submitted by the database authors. Such papers must be published.

Third, published descriptions of full genome sequences typically do not explain in detail the sequence-analysis procedure used for the genome, e.g. exactly what set of analysis programs were applied to each sequence, and how were their outputs combined? Such accounts would again provide default metadata for interpreting the annotation of each genome.

Many of these ideas have been circulating in the bioinformatics community for years. The GSDB project at the National Center for Genome Resources and the GAIA project at the University of Pennsylvania have addressed some of these issues by creating database schemas that represent extensive metadata about sequence annotations, but their ideas have not been adopted by the other public sequence DBs. By allowing these problems to fester we risk a state of affairs where the scientific community loses all faith in the annotations within the public databases, and where the databases have become so large that they cannot be revised within an acceptable time frame.

The following recommendations address these problems:

- The public sequence DBs should develop next-generation schemas that encode metadata about sequence annotations. Significant value can be obtained quickly from relatively simple schema extensions.
- The public sequence DBs should thoroughly document their operating procedures, such as annotation strategies and update policies.
- Research is required to estimate the current error rates of the sequence DBs.
- Research is required to estimate the error rate of functional annotation by different methods of computational sequence analysis.
- A sequence-analysis benchmarking suite should be developed to allow systematic evaluation of automated programs that perform functional annotation.

Peter D. Karp
pkarp@PangeaSystems.Com

Sequences and topology

Deriving biological knowledge from genomic sequences

Editorial overview

Peer Bork* and David Eisenberg†

Addresses

*European Molecular Biology Laboratory, Meyerhofstrasse 1, 102209, D-69012 Heidelberg, Germany;
e-mail: bork@embl-heidelberg.de

†UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Molecular Biology Institute, University of California Los Angeles, Box 951570, Los Angeles, CA 90095-1570, USA;
e-mail: david@mbi.ucla.edu

Current Opinion in Structural Biology 1998, 8:331–332

http://biomednet.com/elecref/0959440X00800331

The nine reviews in this section chart new methods for understanding the biological messages of genome sequences. The accelerating rate at which these sequences are being determined has created a demand for informative analytical methods. The accumulation of new data does not in itself lead to increased knowledge. Rather, it challenges us to improve methods for the filtering and processing of sequences to identify the subtle signals therein. This need is heightened by the advent of sequences of entire genomes; these allow qualitatively new features to be detected and open new views on the evolution of genetic material. The initial progress of this emerging science of functional genomics is impressive and is documented in this set of reviews.

Fortunately, one of the first observations to emerge from comparative genome analysis is the robustness of genetic material that has undergone rearrangement. It may be shuffled, horizontally transferred and disrupted, but nevertheless it often maintains its functionality in different organisms. One of the biological themes seems to be 'modularity', which shows up in noncoding DNA, as well as within the genes, and is also manifest in the three-dimensional structures of their products.

Modularity in DNA is created by duplication events followed by modifications, leading to repetitive segments of DNA. Jerzy Jurka (pp 333–337) reviews the evolution of the repetitive transposable elements that comprise a considerable fraction of the total DNA in eukaryotic genomes. Classification and improved detection is essential for genome annotation and also for cleaning expressed sequence-tag databases. Jurka emphasizes that the previous view, that these repeats are merely selfish elements, needs to be expanded. Also, whereas most of the current applications treat repeats only as 'waste' for the reduction of search space, the repeats seem to have diverse roles in the genome that can be exploited in a wide range of applications, ranging from population studies to mapping and genomic engineering.

Detection and analysis of repeats is also a challenge at the protein level. Jaap Heringa (pp 338–345) reviews the shift in focus during the past year from repeats at the protein domain level to much shorter fragments that are associated with protein malfunction and genetic diseases. At both the domain level and the subdomain level, the relationship between sequence repeats and three-dimensional structure remains a puzzle.

After the detection of repeats, it is crucial to identify the genes in the genomes. Christopher Burge and Samuel Karlin (pp 346–354) review the recent progress in method development, and also point out future directions. The problem of finding genes (particularly in eukaryotes) is far from solved. No wonder, because various weak translational, transcriptional and splicing signals in the DNA have to be identified and combined with experimental information, such as from expressed sequence tags and trapped exons.

Identification of genes is essential, but their full value comes only with their functional and structural annotation. Using the first complete prokaryotic genomes, Eugene Koonin and colleagues (pp 355–363) discuss important aspects of this annotation process, such as the identification of orthologs and the assignment of folds and catalytic activities. The power of comparative sequence analysis, well known at the level of individual proteins, is now also found at the genome level.

There is still much, however, that is not evident from sequence. Genetic mechanisms can cause modifications of sequence (such as circular permutations, domain insertions and secondary-structure rearrangements) that are beyond the limits of detection of current sequence analysis methods. Robert Russell and Christ Ponting (pp 364–371) summarize cases that can be deciphered only by the analysis of protein topology. Their review emphasizes a general point: in many cases, only structural information can illuminate some of the phenomena that hamper sequence analysis.

Structural knowledge can increase the sensitivity of sequence searches. Liisa Holm (pp 372–379) shows how one can exploit superposition of three-dimensional structures for the unification of protein sequence families and the detection of remote homologues. Yet structural similarity does not lead to iron-clad functional predictions

because the same fold can support numerous functions. This is illustrated by the examples that Alexey Murzin (pp 380–387) presents. These examples also show how a wealth of structural data can be correlated in the light of protein evolution.

The complexity of the course of evolution adds complications to genomic analysis. Structural similarity does not necessarily mean a common evolutionary origin and homologous sequences may evolve into different folds (according to current classification schemes). A single function can be found on similar structural scaffolds, so there are numerous examples of parallel evolution towards a similar functionality, even based on extremely different folds. This adds complexity to sequence annotation, as most of the current knowledge on sequenced genomes (particularly beyond the well characterized yeast and *Escherichia coli* genomes) comes from functional inference via homology searches. Thus we can never be sure that a detected homologue has exactly the same function in different genomes. On the other hand, when we hunt for a particular function in a genome, it is always possible that an unrelated protein has acquired this particular function.

A first step towards clarifying such problems will be reliable functional annotation that discriminates between *in vivo*, *in vitro* and (homology) derived data. Clarification also requires, where possible, a structure-based annotation of functional features. At the start, we need to ask what kind of features can and should be derived and described for each sequence. Functional classifications are essential if we want to describe metabolism and, ultimately, phenotypes. Monica Riley (pp 388–392) summarizes many of the problems in function classification, including semantics, hierarchies and inconsistencies. It is important to reach a consistent annotation level, but will we ever achieve annotation that is both reasonably complete and computer-readable? Function always depends on the context and yet only molecular features can be deduced directly from sequence. Some information comes from the availability of entire genomes: for example, the absence of genes and/or functions can be included in predictions.

Today, what we predict from sequences is at best fragmentary and qualitative, for example, the presence or absence of a certain gene or structure or function or pathway. This is not enough to describe cellular processes. Fortunately, there are experimental tools of growing power for the support and extension of genome predictions, such as direct measures of gene expression and protein interaction. One of the leading techniques is mass spectrometry. Bernhard Küster and Matthias Mann (pp 393–400) describe how mass spectrometry can be used to sequence and identify proteins that have post-translational modifications, even though some cannot yet be predicted from sequence.

Although sequence and structure space is not infinite, we will probably never be able to explore them completely (consider, for example, the extinction of species with their genetic material and the rapid modification of virus sequences). With model genomes from evolutionarily distant species becoming available, however, we can make a start at this exploration for humans and other living organisms. In this endeavor, the methods for analysis and annotation that are being developed today will be of the utmost importance in future attempts to bridge the genotype and phenotype of organisms.

X. Related Proceedings Appendix

This Appendix has no copies of decisions in related proceedings because there are no related proceedings to this Appeal.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/052,664	01/17/2002	Paul David Cannon	ROCH-001DIV DC	4008

24372 7590 08/26/2005

ROCHE PALO ALTO LLC
PATENT LAW DEPT. M/S A2-250
3431 HILLVIEW AVENUE
PALO ALTO, CA 94304

RECEIVED
AUG 30 2005
ROCHE PALO ALTO
PATENT DEPT.

EXAMINER	
BASI, NIRMAL SINGH	

ART UNIT	PAPER NUMBER
1646	DC

DATE MAILED: 08/26/2005

R0058C-Div
9/26/05 Due
3/26/06 Final

Please find below and/or attached an Office communication concerning this application or proceeding.

Notification of Non-Compliant Appeal Brief (37 CFR 41.37)	Application No.	Applicant(s)	
	10/052,664	CANNON ET AL.	
	Examiner	Art Unit	
	Nirmal S. Basi	1646	

--The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

The Appeal Brief filed on 20 October 2004 is defective for failure to comply with one or more provisions of 37 CFR 41.37.

To avoid dismissal of the appeal, applicant must file a complete new brief in compliance with 37 CFR 41.37 within **ONE MONTH or THIRTY DAYS** from the mailing date of this Notification, whichever is longer. **EXTENSIONS OF THIS TIME PERIOD MAY BE GRANTED UNDER 37 CFR 1.136.**

1. ☒ The brief does not contain the items required under 37 CFR 41.37(c), or the items are not under the proper heading or in the proper order.

2. ☐ The brief does not contain a statement of the status of all claims, (e.g., rejected, allowed or confirmed, withdrawn, objected to, canceled), or does not identify the appealed claims (37 CFR 41.37(c)(1)(iii)).

3. ☐ At least one amendment has been filed subsequent to the final rejection, and the brief does not contain a statement of the status of each such amendment (37 CFR 41.37(c)(1)(iv)).

4. ☒ (a) The brief does not contain a concise explanation of the subject matter defined in each of the independent claims involved in the appeal, referring to the specification by page and line number and to the drawings, if any, by reference characters; and/or (b) the brief fails to: (1) identify, for each independent claim involved in the appeal and for each dependent claim argued separately, every means plus function and step plus function under 35 U.S.C. 112, sixth paragraph, and/or (2) set forth the structure, material, or acts described in the specification as corresponding to each claimed function with reference to the specification by page and line number, and to the drawings, if any, by reference characters (37 CFR 41.37(c)(1)(v)).

5. ☒ The brief does not contain a concise statement of each ground of rejection presented for review (37 CFR 41.37(c)(1)(vi))

6. ☐ The brief does not present an argument under a separate heading for each ground of rejection on appeal (37 CFR 41.37(c)(1)(vii)).

7. ☒ The brief does not contain a correct copy of the appealed claims as an appendix thereto (37 CFR 41.37(c)(1)(viii)).

8. ☒ The brief does not contain copies of the evidence submitted under 37 CFR 1.130, 1.131, or 1.132 or of any other evidence entered by the examiner and relied upon by appellant in the appeal, along with a statement setting forth where in the record that evidence was entered by the examiner, as an appendix thereto (37 CFR 41.37(c)(1)(ix)).

9. ☐ The brief does not contain copies of the decisions rendered by a court or the Board in the proceeding identified in the Related Appeals and Interferences section of the brief as an appendix thereto (37 CFR 41.37(c)(1)(x)).

10. ☒ Other (including any explanation in support of the above items):

See Continuation Sheet.

ANTHONY C. CAPUTA
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 1600

Continuation of 10. Other (including any explanation in support of the above items): A review by the Board of Patent Appeals and Interferences of the application reveals that the following sections are missing from the Appeal Brief:
(1) "Summary of claimed subject matter" as set forth in 37 CFR 41.37(c)(1)(v);
(2) "Grounds of rejection to be reviewed on appeal" as set forth in 37 CFR 41 .37(c)(1)(vi) (replaced "Issues for Review" and "Grouping of Claims");
(3) "Claims appendix," as set forth in 37 CFR 41 .37(c)(1)(viii) (replaced "Appendix");
(4) "Evidence appendix," as set forth in 37 CFR 41 .37(c)(1)(ix); and
(5) "Related proceedings appendix," as set forth in 37 CFR 41.37(c)(1)(x).
Accordingly, the Appeal Brief filed on October 20, 2004 does not comply with the new rules under 37 CFR 41.37(c). It is required that a substitute Appeal Brief be submitted that is in compliance with 37 CFR 41.37(c). For more information on the Board's new rules, please see the web page entitled "More Information on the Rules of Practice Before the BPAI," Final Rule at:
http://www.uspto.gov/web/offices/dcom/bpai/fr2004/moreinfo.html